

The eukaryote chromosome

Editors W. J. Peacock and R. D. Brock



This book was published by ANU Press between 1965–1991.

This republication is part of the digitisation project being carried out by Scholarly Information Services/Library and ANU Press.

This project aims to make past scholarly works published by The Australian National University available to a global audience under its open-access policy.

The eukaryote chromosome

PLEASE RETURN TO :-
EDITORIAL DEPARTMENT
AUSTRALIAN NATIONAL UNIVERSITY

RECOMMENDED RETAIL PRICE 15.95
PUBLICATION DATE 22-8-75

The eukaryote chromosome

Editors W. J. Peacock and R. D. Brock

**Papers presented at a conference held under the
auspices of the U.S./Australia Science Agreement**

Australian National University Press Canberra 1975

First published in Australia 1975

Printed in Australia for the Australian National University Press, Canberra

©The Australian National University 1975

This book is copyright. Apart from any fair dealing for the purpose of private study, research, criticism, or review, as permitted under the Copyright Act, no part may be reproduced by any process without written permission. Inquiries should be made to the publisher.

National Library of Australia
Cataloguing-in-Publication entry

The eukaryote chromosome: papers presented at a conference held under the auspices of the U.S./Australia Science agreement/Editors: W. J. Peacock and R. D. Brock. — Canberra: Australian National University Press, 1975.

Conference held in Canberra, 1973.

Index.

Includes bibliographies.

ISBN 0 7081 0104 6.

1. Chromosomes — Congresses. 2. Eukaryotic cells — Congresses. I. Peacock, William James, ed. II. Brock, Richard Donald, joint ed.

547.8732

North, South, and Central America:

International Scholarly Book Services, Inc., Portland, Oregon. **Southeast Asia:** Angus & Robertson (S.E. Asia) Pty Ltd, Singapore. **Japan:** United Publishers Services Ltd, Tokyo.

This conference was sponsored by the United States Academy of Science/Australian Department of Science.

CONTENTS

List of Contributors viii

The arrangement of DNA sequences in the chromosomes of *Drosophila melanogaster*

David S. Hogness, Pieter C. Wensink, David M. Glover, Raymond L. White, David J. Finnegan and John E. Donelson 1

Sequences of highly repeated DNA in *Drosophila melanogaster*

D. L. Brutlag and W. J. Peacock 35

Arrangement of the highly repeated DNA of *Drosophila melanogaster*

E. S. Goldring, D. L. Brutlag and W. J. Peacock 47

The organization of eukaryotic DNA and the sequence environment of structural genes

Eric H. Davidson, Barbara R. Hough, Michael J. Smith, Dale E. Graham, William H. Klein, Glenn A. Galau, Margaret E. Chamberlin, and Roy J. Britten 61

An analysis of *Drosophila* DNA by restriction endonucleases

C. A. Thomas Jr, D. H. Hamer, R. E. Pyeritz, G. P. Holmquist, K. Shastri and S. Manteuil 87

Circular DNA in the yeast *Saccharomyces cerevisiae* in relation to informosomal DNA and mitochondrial genetics

G. D. Clark-Walker and George L. Gabor Miklos 111

Chromosome structure: evidence against the unineme hypothesis

Judith H. Ford 123

Ribosomal gene amplification in oocytes and in primitive eukaryotes

Joseph G. Gall 135

Differential replication of rDNA during insect development

J. A. Thomson 149

In vivo purification of mitochondrial genes in petite mutants of yeast: the genetic and biochemical characterization of mutants enriched in mitochondrial ribosomal RNA cistrons

Phillip Nagley, P. L. Molloy, H. B. Lukins and Anthony W. Linnane 155

Genes and chromomeres of *Drosophila*

Burke H. Judd 169

The structure and regulation of eukaryotic genomes approached by studying HnRNA

James E. Darnell Jr 185

Oligothymidylic acid tracts in mouse DNA

Jonathan N. Mansbridge, George D. Birnie, Steve Humphries, Som Atisoontornkul and John Paul 199

Properties of the mRNA for feather keratin: mRNA transcribed from a family of repeated genes

David J. Kemp, Ian D. Walker, Geoffrey A. Partington and George E. Rogers 205

DNA sequences from which double-stranded RNA is transcribed

H. Naora, J. M. Whitelam, R. P. Monckton and R. E. H. Wettenhall 217

Evidence for mRNA of higher plants

T. J. V. Higgins and P. B. Goodwin 233

A model of chromatin structure

Roger D. Kornberg 245

Origins of chromatin acidic proteins of avian erythroid cells

R. W. H. Harlow and J. R. E. Wells 255

Chromosomal proteins and aspects of the transcription process in eukaryotes

Ronald J. Hill 269

Zygotene and pachytene-labeled sequences in the meiotic organization of chromosomes

Yasuo Hotta and Herbert Stern 283

Regulation of genetic recombination in *Neurospora crassa*

D. G. Catcheside 301

The synaptonemal complex in the course of meiotic chromosome pairing and recombination

C. B. Gillies 313

Replication initiation in yeast chromosomes

Wylie Burke, Carol Shaw Newlon and Walton L. Fangman 327

Replication of polyoma DNA in isolated nuclei

Roger F. Martin 345

Chromosome functional organization determined by replication mapping

Chev Kidson, Phillip C. Chen, Tony Edwards and Jonathan Mansbridge 355

Control of DNA synthesis in somatic cell hybrids

Jennifer A. Marshall Graves 367

Sex chromosome evolution and activity in mammals, particularly kangaroos

D. W. Cooper, P. G. Johnston, C. E. Murtagh and J. L. VandeBerg 381

Heterochromatin and timing of DNA replication in morabine grasshoppers

G. C. Webb and M. J. D. White 395

Studies of the molecular anatomy of the SV40 viral genome

Paul Berg, Janet E. Mertz and John Carbon 409

Mapping of mutational alterations in DNA with S1 nuclease: the location of deletions, insertions and temperature-sensitive mutations in SV40

Thomas E. Shenk, Carl Rhodes, Peter W. J. Rigby and Paul Berg* 431

The transfer and expression (transgenosis) of foreign genes in plant cells, reality and potential

Colin H. Doy 447

Characteristics of bacterial gene expression in human fibroblasts

Carl R. Merrill, Mark R. Geier and Barry G. Rolfe 459

Resumé of the conference

Dan L. Lindsley 473

Index 495

*This paper has also been published in Cold Spring Harbor Symposia on Quantitative Biology, Vol. 39, 1975.

Contributors

Dr. Som Atisoontornkul,
Department of Biochemistry,
University of Queensland,
ST. LUCIA, QLD, 4067, AUSTRALIA.

Dr. Paul Berg,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. George D. Birnie,
The Beatson Institute for Cancer Research,
132 Hill Street,
GLASGOW, G3 6UD, SCOTLAND.

Dr. Roy J. Britten,
Kerckhoff Marine Laboratory,
California Institute of Technology,
101 Dahlia Street,
CORONA DEL MAR, Calif. 92625, U.S.A.

and

Carnegie Institute of Washington,
5241 Broad Branch Road, N.W.,
WASHINGTON, D.C., 20015, U.S.A.

Dr. D.L. Brutlag,
CSIRO Division of Plant Industry,
P.O. Box 1600,
CANBERRA CITY, A.C.T. 2601, AUSTRALIA.

Present address:
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. Wylie Burke,
Department of Genetics,
University of Washington,
SEATTLE, Wash., 98195, U.S.A.

Dr. John Carbon,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. D.G. Catcheside,
Department of Genetics,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T. 2600, AUSTRALIA.

Dr. Margaret E. Chamberlin,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. Phillip C. Chen,
Department of Biochemistry,
University of Queensland,
ST. LUCIA, QLD, 4067, AUSTRALIA.

Dr. G.D. Clark-Walker,
Department of Developmental Biology,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T. 2600, AUSTRALIA.

Dr. D.W. Cooper,
School of Biological Sciences,
Macquarie University,
NORTH RYDE, N.S.W., 2113, AUSTRALIA.

Dr. James E. Darnell Jr.,
Department of Biological Sciences,
Columbia University,
NEW YORK, N.Y. 10027, U.S.A.

Dr. Eric H. Davidson,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. John E. Donelson,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Present address:
School of Medicine,
University of Iowa,
IOWA CITY, IOWA, 52240, U.S.A.

Dr. Colin H. Doy,
Department of Genetics,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. Tony Edwards,
Department of Biochemistry,
University of Queensland,
ST. LUCIA, QLD, 4067, AUSTRALIA.

Dr. Walton L. Fangman,
Department of Genetics,
University of Washington,
SEATTLE, Wash., 98195, U.S.A.

Dr. David J. Finnegan,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. Judith H. Ford,
Medical Research Department,
Kanematsu Memorial Institute,
Sydney Hospital,
Macquarie Street,
SYDNEY, N.S.W., 2000, AUSTRALIA.

Dr. Joseph G. Gall,
Department of Biology,
Yale University,
NEW HAVEN, Conn., 06520, U.S.A.

Dr. Glenn A. Galau,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. Mark R. Geier,
Laboratory of General and Comparative Biochemistry,
National Institute of Mental Health,
BETHESDA, Md, 20014, U.S.A.

Dr. C.B. Gillies,
School of Biological Sciences,
University of Sydney,
SYDNEY, N.S.W. 2000, AUSTRALIA.

Dr. David M. Glover,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. E.S. Goldring,
CSIRO Division of Plant Industry,
P.O. Box 1600,
CANBERRA CITY, A.C.T., 2601, AUSTRALIA.

Present address:
Department of Biochemistry,
School of Medicine,
University of Papua New Guinea,
BOROKO, P.N.G.

Dr. P.B. Goodwin,
Department of Developmental Biology,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T. 2600, AUSTRALIA.

Dr. Dale E. Graham,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Present address:

Department of Biological Sciences,
Purdue University,
LAFAYETTE, INDIANA, 47907, U.S.A.

Dr. Jennifer A. Marshall Graves,
Department of Genetics and Human Variation,
La Trobe University,
BUNDOORA, VIC., 3083, AUSTRALIA.

Dr. D.H. Hamer,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. R.W.H. Harlow,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A., 5001, AUSTRALIA.

Dr. T.J.V. Higgins,
CSIRO Division of Plant Industry,
P.O. Box 1600,
CANBERRA CITY, A.C.T., 2601, AUSTRALIA.

Dr. Ronald J. Hill,
CSIRO Division of Animal Genetics,
P.O. Box 90,
EPPING, N.S.W., 2121, AUSTRALIA.

Dr. David S. Hogness,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. G.P. Holmquist,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. Yasuo Hotta,
Department of Biology,
University of California, San Diego,
LA JOLLA, Calif., 92037, U.S.A.

Dr. Barbara R. Hough,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. Steve Humphries,
The Beatson Institute for Cancer Research,
132 Hill Street,
GLASGOW, G3 6UD, SCOTLAND.

Dr. P.G. Johnston,
School of Biological Sciences,
Macquarie University,
NORTH RYDE, N.S.W., 2113, AUSTRALIA.

Dr. Burke H. Judd,
Department of Zoology,
University of Texas at Austin,
AUSTIN, TEXAS, 78712, U.S.A.

Dr. David J. Kemp,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A., 5001, AUSTRALIA.

Dr. Chev Kidson,
Department of Biochemistry,
University of Queensland,
ST. LUCIA, QLD, 4067, AUSTRALIA.

Dr. William H. Klein,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. Roger D. Kornberg,
MRC Laboratory of Molecular Biology,
Hills Road,
CAMBRIDGE, CB2 2QH, ENGLAND.

Dr. Dan L. Lindsley,
Department of Biology,
University of California, San Diego,
LA JOLLA, Calif., 92037, U.S.A.

Dr. Anthony W. Linnane,
Department of Biochemistry,
Monash University,
CLAYTON, VIC., 3168, AUSTRALIA.

Dr. H.B. Lukins,
Department of Biochemistry,
Monash University,
CLAYTON, VIC., 3168, AUSTRALIA.

Dr. Jonathan N. Mansbridge,
Department of Biochemistry,
University of Queensland,
ST. LUCIA, QLD., 4067, AUSTRALIA.

Dr. S. Manteuil,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. Roger F. Martin,
Biological Research Unit,
Cancer Institute,
278 William Street,
MELBOURNE, VIC., 3000, AUSTRALIA.

Dr. Janet E. Mertz,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. Carl R. Merrill,
Laboratory of General and Comparative Biochemistry,
National Institute of Mental Health,
BETHESDA, Md, 20014, U.S.A.

Dr. George L. Gabor Miklos,
Department of Genetics,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. P.L. Molloy,
Department of Biochemistry,
Monash University,
CLAYTON, VIC., 3168, AUSTRALIA.

Dr. R.P. Monckton,
Molecular Biology Unit,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. C.E. Murtagh,
School of Biological Sciences,
Macquarie University,
NORTH RYDE, N.S.W., 2113, AUSTRALIA.

Dr. Phillip Nagley,
Department of Biochemistry,
Monash University,
CLAYTON, VIC., 3168, AUSTRALIA.

Dr. H. Naora,
Molecular Biology Unit,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. Carol Shaw Newlon,
Department of Genetics,
University of Washington,
SEATTLE, Wash., 98195, U.S.A.

Present address:

Department of Zoology,
University of Iowa,
IOWA CITY, IOWA, OHIO, 52242, U.S.A.

Dr. Geoffrey A. Partington,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A., 5001, AUSTRALIA.

Present address:

Institute of Neurology,
Department of Neurochemistry,
8 Exchange Court,
LONDON, WC2, ENGLAND.

Dr. John Paul,
The Beatson Institute for Cancer Research,
132 Hill Street,
GLASGOW, G3 6UD, SCOTLAND.

Dr. W.J. Peacock,
CSIRO Division of Plant Industry,
P.O. Box 1600,
CANBERRA CITY, A.C.T., 2601, AUSTRALIA.

Dr. R.E. Pyeritz,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. Carl Rhodes,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. Peter W.J. Rigby,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. George E. Rogers,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A., 5001, AUSTRALIA.

Dr. Barry G. Rolfe,
Department of Genetics,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. K. Shastri,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. Thomas E. Shenk,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Dr. Michael J. Smith,
Division of Biology,
California Institute of Technology,
PASADENA, Calif., 91109, U.S.A.

Dr. Herbert Stern,
Department of Biology,
University of California, San Diego,
LA JOLLA, Calif., 92037, U.S.A.

Dr. C.A. Thomas, Jr,
Department of Biological Chemistry,
Harvard Medical School,
BOSTON, Mass., 02115, U.S.A.

Dr. J.A. Thomson,
Genetics Department,
University of Melbourne,
PARKVILLE, VIC., 3052, AUSTRALIA.

Present address:
CSIRO Division of Plant Industry,
P.O. Box 1600,
CANBERRA CITY, A.C.T., 2601, AUSTRALIA.

Dr. J.L. VandeBerg,
School of Biological Sciences,
Macquarie University,
NORTH RYDE, N.S.W., 2113, AUSTRALIA.

Dr. Ian D. Walker,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A. 5001, AUSTRALIA.

Present address:

MRC Laboratory of Molecular Biology,
University Postgraduate Medical School,
CAMBRIDGE, CB2 2QH, ENGLAND.

Dr. G.C. Webb,
Department of Population Biology,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. Pieter G. Wensink,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Present address:
Rosenstiel Center,
Brandeis University,
WALTHAM, Mass., 02154, U.S.A.

Dr. J.R.E. Wells,
Department of Biochemistry,
University of Adelaide,
ADELAIDE, S.A., 5001, AUSTRALIA.

Dr. R.E.H. Wettenhall,
Molecular Biology Unit,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

Dr. M.J.D. White,
Department of Genetics,
University of Melbourne,
PARKVILLE, VIC., 3052, AUSTRALIA.

Dr. Raymond L. White,
Department of Biochemistry,
Stanford University School of Medicine,
STANFORD, Calif., 94305, U.S.A.

Ms. J.M. Whitelam,
Molecular Biology Unit,
Research School of Biological Sciences,
Australian National University,
CANBERRA, A.C.T., 2600, AUSTRALIA.

The arrangement of DNA sequences in the chromosomes of *Drosophila melanogaster*

David S. Hogness, Pieter C. Wensink,
David M. Glover, Raymond L. White,
David J. Finnegan and John E.
Donelson

Numerous models have been proposed to explain why only one or a few cistrons are detected per *Drosophila melanogaster* chromomere, although these chromomeres contain sufficient DNA to code for a much larger number of polypeptides. Since these models can be distinguished according to the topography that each imposes on the repetitive and non-repetitive sequences, and on the transcribed and translated sequences, we have developed a system for mapping these sequences within and among the chromomeres. It is based on the isolation of segments of chromosomal DNA in homogeneous form. This is accomplished by cloning hybrid molecules consisting of a segment of *D. melanogaster* DNA inserted into a plasmid DNA which can replicate autonomously in *Escherichia coli*. Both the poly(dA) : poly(dT) and the EcoR1-ligase methods have been used to construct the hybrid plasmids. Nine hybrids have been isolated and the chromosomal segments analyzed for sequence repetition within the segment and within the genome. Their sequences have also been mapped by *in situ* hybridization to polytene chromosomes. Six segments belong to a single group defined by little or no repetition within the genome and localization to a single chromomeric region. The characteristics of this group are not compatible with tandem repetition models for the chromomere. The remaining three segments contain sequences repeated 30 to 400 times per genome, although one also contains non-repetitive sequences. All contain sequences found in the centromeric heterochromatin. One contains sequences which map at 15 different chromomeric regions in addition to the chromocenter. The third contains the gene for 18-28S rRNA.

THE CHROMOMERE ANOMALY

The haploid genome of *Drosophila melanogaster* contains 165,000 kb of DNA (Rasch et al., 1971; Rudkin, 1972) (kb, kilobases, is a unit of length equal to 1,000 bases or base pairs in single- or double-stranded nucleic acids, respectively). This DNA appears to be distributed among only four linear double-stranded molecules, one for each chromosome in a haploid set (Kavenoff and Zimm, 1973). The portion of these molecules that is contained in the euchromatic arms of the chromosome includes c. 130,000 kb (Rudkin, 1972), and is divisible into approximately 5,000 segments, which we shall call chromomeric units. These units are recognized by the bands they form in polytene chromosomes, and are individually defined as the segment of the chromosomal DNA molecule which includes the sequence of base pairs present in a band plus an adjacent interband.

A major problem concerning the arrangement of genetic information in eukaryotic chromosomes is posed by two contrasting properties of these chromomeric units: namely, the amount of DNA and the number of cistrons that they contain. The mean length of the units calculated from the above data is 26 kb, i.e., 130,000 kb/5,000 units. This value depends on the assumption that the euchromatic portion of the chromosomal DNA molecules can be neatly divided into 5,000 segments, each of which produces a band-interband during formation of the polytene arms. It does not depend on most other particulars of polytene structure, as for example, whether the segment corresponding to one band-interband is replicated more or less than another such segment during polytene formation.

The determination of the length of individual units does, however, depend on a particular model for polytene structure. This model is given in Figure 1 and provides the most satisfactory explanation of the data on polytene chromosomes (for reviews, see Beermann, 1972; Rudkin, 1972). The polytene arms are imagined to consist of many identical chromatin fibers (chromatids) that are coordinately aggregated so that the structure of a single fiber is laterally amplified to produce the visible band pattern. The fiber is therefore conceived to fold or coil into compacted regions, or chromomeres, that alternate with regions of no folding, or interchromomeres. The length of a chromomeric unit should then equal the DNA content of the respective band-interband divided by the number of fibers, which is determined from the total DNA content of the polytene arms. Using this method, Rudkin (1961, 1972) determined the lengths of 100 individual units in *D. melanogaster* and obtained a mean of 32 kb and a range of 4.5 to c. 550 kb. (Our calculation of Rudkin's (1961) original data on 60 bands yields the same mean, a range of 7 to 92 kb, and a mode and median of 13 to 26 kb, respectively; see Blumenthal et al., 1974). Beermann (1972) estimates a range of about 5 to 100 kb.

Both determinations of the mean length indicate that chromomeric units of this size contain enough DNA to code for 24-30 polypeptides of average size, i.e., c. 40,000 daltons (Laird, 1971). And if we accept the model in Figure 1, then the larger units could code for 100 or more polypeptides of this size. Were these segments derived from prokaryotes, we should expect that each would contain a similar number of cistrons. However, the cytogenetic experiments of Judd et al. (1972) and of Hochmann (1974) indicate that the number of cistrons per chromomeric unit is considerably less than this prokaryotic expectation. Indeed they suggest that the number may be only one, regardless of the length of the unit (see Judd, this volume, for further discussion of this point). It is this high DNA to cistron ratio that we call the chromomere anomaly.

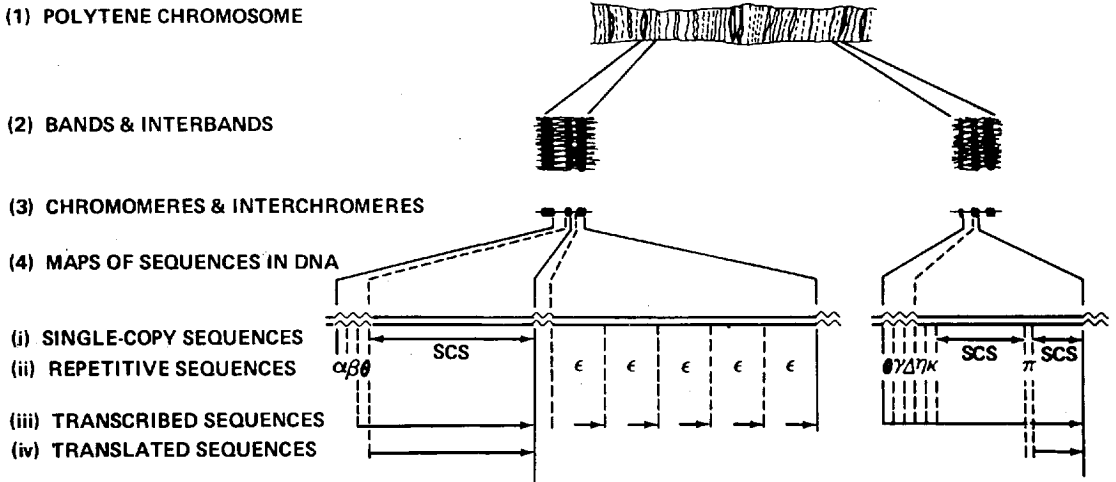


Figure 1. Sequence arrangements in chromomeric units. The sequence topographies of the three chromomeric units shown in this figure illustrate models that are discussed in the text.

MODELS

A growing number of models or explanations for this anomaly have appeared in the literature (for reviews, see Beermann, 1972; Bishop, 1974). They range from the suggestion of O'Brien (1973) that most cistrons escape detection by the applied mutant screens, to the inference from Prescott and Murti's (1974) work with *Styloichia* that the large majority of chromomeric DNA is not transcribed, translated, or used to regulate these activities. What we wish to emphasize here is that most of the models can be distinguished by the topography that each imposes on the repetitive and single-copy sequences, and on the transcribed and translated sequences within and among the chromomeric units.

Thus, models based on the tandem repetition of a sequence such as ϵ in Figure 1 (Callan, 1967; Thomas, 1970; Lee and Thomas, 1973) yield a different arrangement of repetitive sequences than do models in which these sequences provide a regulatory function (Britten and Davidson, 1969; Georgiev, 1969; Crick, 1971; Paul et al., 1974). In the latter case, chromomeric units that respond to the same regulatory molecule may be expected to contain the same sequence, e.g., the sequence θ given in the left- and right-hand chromomeric units of Figure 1. The number of times θ is repeated in the genome will depend on the number of chromomeric units belonging to the set controlled by the θ -recognizing regulatory molecule, and these may be widely interspersed within the genome. If a given chromomeric unit

belongs to several different regulatory sets, then we expect that it will contain several different repetitive sequences (represented by the other Greek letters in Figure 1), each capable of being recognized by the regulatory molecule that defines the set.

Similarly, many models can be differentiated according to which sequences in the unit are transcribed and which are translated. In certain models (e.g., see Darnell, this volume), most of the unit is transcribed to produce long HnRNA molecules of which only a small fraction is retained in the mRNA for translation (right-hand unit, Figure 1). Or, if many cistrons in a chromomeric unit have passed undetected through the mutant screens then the translated sequences should also occupy a large fraction of the unit (left-hand unit, Figure 1).

EXPERIMENTAL PLAN

Put in its simplest form, the plan is to map these relevant sequences in a representative set of chromosomal segments from *D. melanogaster*, and thereby reduce the number of allowable models to a more manageable set. A prerequisite for this mapping is the isolation of homogeneous populations of individual segments to which we can apply the techniques that have been developed for mapping sequences in viral DNAs (for examples see Berg et al. and Shenk et al., this volume). Figure 2 illustrates the general plan we formulated to achieve this isolation. In the first step, chromosomal segments are inserted into the circular DNA of a plasmid that can replicate autonomously in *Escherichia coli*. The resulting heterogeneous population of hybrid DNAs is then used to transform *E. coli* with respect to a function that is provided by the parent plasmid. Homogeneity is achieved by isolating clones of transformants that result from the infection of one bacterium by a single hybrid molecule; i.e., autonomous replication of this molecule in the expanding clone provides the desired homogeneous population of DNA molecules that are isolated from the bacteria in the last step of the procedure.

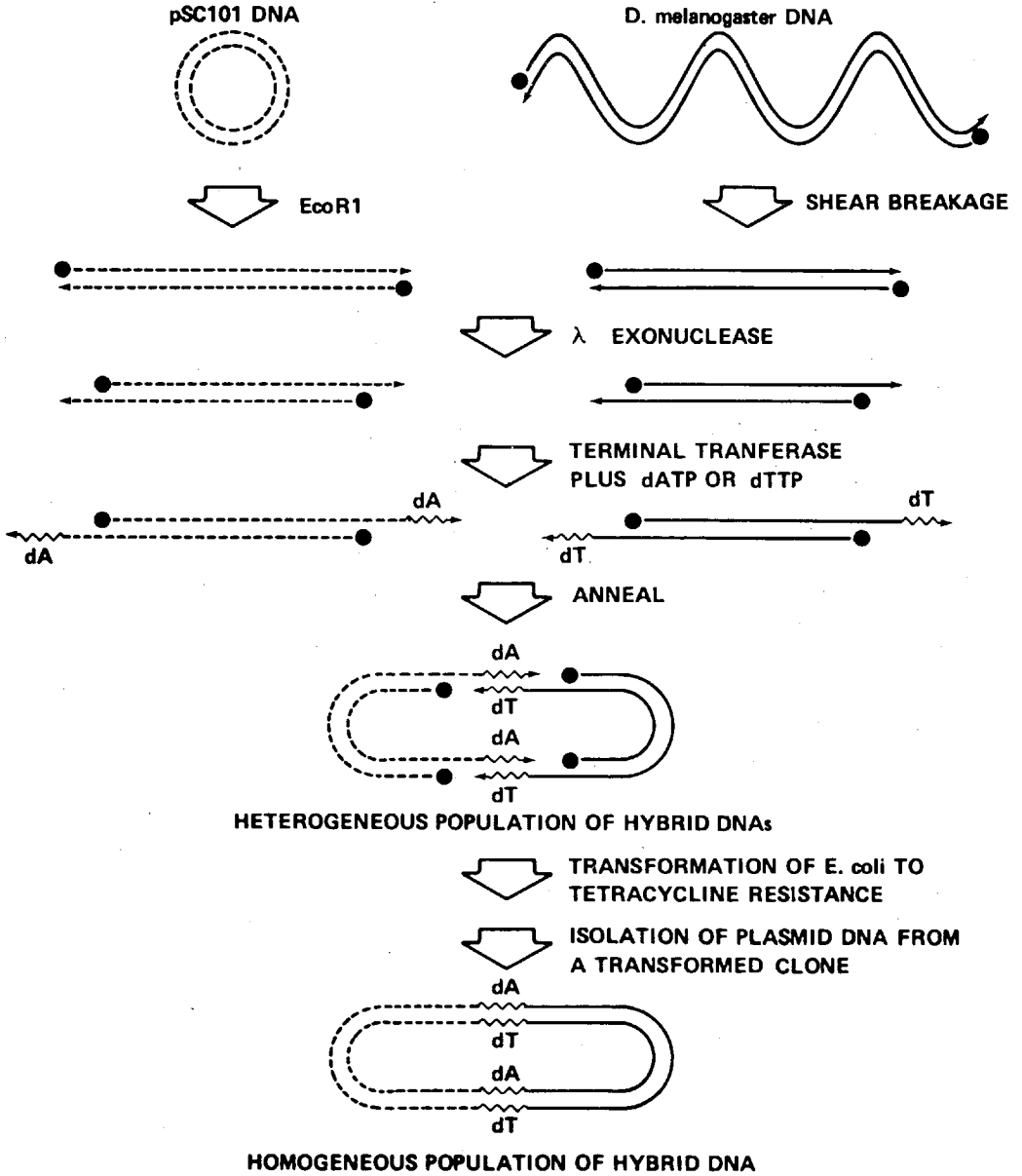
The insertion is accomplished by introducing a double-stranded break into the circular plasmid DNA and then joining each of the resultant termini to the two termini of a chromosomal segment. We have used two methods of joining the DNA segments to form the plasmid-*D. melanogaster* hybrids (abbreviated by pDm). The poly(dA):poly(dT) method shown in Figure 2 was adapted from the technique developed by Jackson et al. (1972) and by Lobban and Kaiser (1973). It consists of adding short poly(dA) or poly(dT) tails to the 3'-OH termini of the respective DNAs, and then annealing these tails to form the circular pDm molecules. The second, or EcoRI-ligase method for joining two DNA molecules was developed by Mertz and Davis (1972)

and is illustrated in Figure 3. In this case, the tails are on the 5'-termini and are created by cleavage of the DNA with the EcoRI restriction endonuclease at specific sites in double-stranded DNA (Hedgpeth et al. 1972). The tail sequences are identical and complementary. Hence, any tail can anneal with any other tail, and at low temperatures the resulting joint exhibits a sufficient life time to allow *E. coli* ligase to covalently join the strands and thereby stabilize the joint.

Two differences between the joining methods are important for our purposes. The pDm DNAs will provide a reliable means for obtaining sequence maps of the *D. melanogaster* chromosomes only if the *D. melanogaster* segment in the hybrid (abbreviated by Dm) derives from a single continuous segment of the chromosomal DNA. The poly(dA):poly(dT) method is designed to satisfy this provision since direct linkage between molecules from the same source does not occur (Jackson et al., 1972; Lobban and Kaiser, 1972). By contrast, EcoRI-ligase joining is indiscriminate, and *ad hoc* provisions are required to ensure that a given Dm segment does not consist of multiple chromosomal segments linked to each other. The provision that we have employed is a careful sizing of the chromosomal segments and of the resulting hybrids to ensure that the length of the hybrid molecules is the sum of the lengths of one plasmid and one chromosomal segment.

The second difference is that segments to be joined by the poly(dA):poly(dT) method can be created by virtually any mechanism from shear breakage to cleavage by specific endonucleases, whereas the EcoRI-ligase method requires that all termini derive from EcoRI cleavage sites. For purposes of building a representative library of Dm segments, the advantage lies with the poly(dA):poly(dT) method since a random collection of chromosomal segments can be obtained by shear breakage but not by EcoRI cleavage. However, the requirement for EcoRI termini in combination with a specification of the segment length can impose a selection for certain kinds of Dm segments that is useful.

The location of productive insertion sites in a plasmid DNA is also a function of the DNA cleavage mechanisms allowed by the joining method. These sites are primarily limited to those regions where insertion does not interfere with the expression of the plasmid genes required for replication and transformation. Within such productive regions, they are secondarily limited to the sites which can be cleaved to produce the termini required by the joining method. The tetracycline resistance plasmid, pSC101, is ideal in this regard since its single EcoRI cleavage site is in a productive region (Cohen et al., 1973; Morrow et al., 1974), and hence the plasmid



reactant for either joining method can be simply obtained by EcoRI cleavage of the circular monomer (Figures 2 and 3).

By contrast the single EcoRI site in *λdvgaI* (Jackson et al., 1972; Mertz and Davis, 1972), the plasmid we used in our early experiments, is found in a region required for replication (Thomas and Davis, 1975). To get around this difficulty, we inserted

Figure 2. Construction and isolation of hybrid plasmids by the poly(dA):poly(dT) procedure (Wensink et al., 1974). Arrowheads and dots represent the 3' - and 5' - termini of the strands, respectively. The details of the procedure are given by Wensink et al. (1974); and are summarized below. *Drosophila melanogaster* (Oregon R) DNA was isolated from 0 to 15 hour embryonic nuclei according to the method of Schachat and Hogness (1974) with minor modifications, and freed of most satellites by sedimentation in actinomycin D/CsCl gradients (Peacock et al., 1974). It was then sheared to a mean length of 8 kb by stirring (Hogness and Simmons, 1964). An average of 25 nucleotides per 5' - terminus were removed from this DNA and from the linear monomer of pSC101 DNA (obtained by EcoRI cleavage of the circular plasmid; Cohen et al., 1973) with λ exonuclease (Little et al., 1967), and an average of about 40 dT or dA nucleotides per 3' - terminus were then added with calf thymus terminal transferase (Kato et al., 1967) under the conditions of Lobban and Kaiser (1973), except that the temperature was lowered to 25°C. The poly(dT)-*D. melanogaster* DNA (1.2 μ g/ml) was annealed to the poly(dA) - pSC101 DNA (0.6 μ g/ml) in 0.1 M NaCl in 1 mM EDTA, 10 mM Tris-HCl, pH 8.2 (TE) for 10 min at 65°C and then 2 hours at 45°C. Transformation of tetracycline sensitive C600, *hsm*⁻, *hsr*⁻, *gal*⁻ (strain h303; Wensink et al., 1974) to resistance with this DNA was effected by a modification of the method of Mandel and Higa (1970), and occurred with an efficiency approximately two orders of magnitude less than that observed with closed circular pSC101 DNA (10⁻⁶ transformants/molecule). Closed circular pDm and pSC101 DNAs were isolated from lysates (Clewell and Helinski, 1969) by centrifugation in ethidium bromide/CsCl gradients (Radloff et al., 1968). Ethidium bromide and contaminating RNA were removed by passage through a column (diam. = 1 cm) with a 2 cm layer of Dowex AG50 on top of 13 cm of Biogel A15M. The column was equilibrated and the DNA eluted with 0.5 M NaCl in TE. All DNA solutions were dialyzed against TE prior to storage at 4°C.

the chromosomal segment at either one of the two EcoRI sites in the circular dimer, thereby leaving an intact λ *dvgal* genome to provide the replication functions for the hybrid. We abandoned λ *dvgal* in favor of pSC101 because this trick involves the laborious isolation of the linear dimer from the mixture obtained by partial EcoRI cleavage of circular dimers.

The pDm molecules consist of a single chromosomal segment linked to a single pSC101 segment

Molecules constructed by the poly(dA):poly(dT) method (Wensink et al., 1974) When products of the annealing reaction (Figure 2) from each of several experiments were analyzed in the electron microscope, the fraction of molecules exhibiting the expected circular structure fell in the range 16 \pm 9%. This fraction dropped to \leq 1% in control experiments where either pSC101 or the *D. melanogaster* DNA was left out of the annealing mixture. The control data are in accord with the previous finding (Jackson et al., 1972; Lobban and Kaiser, 1972) that linkage between tails of the same kind does not occur. Circular hybrids formed by this method should therefore consist of

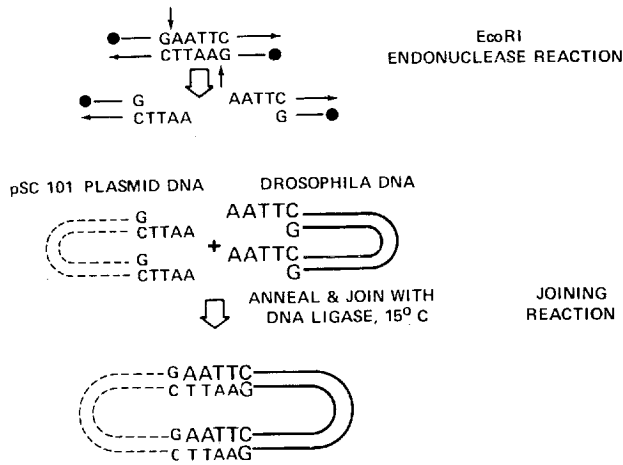


Figure 3. Construction of hybrid plasmids by the EcoRI - ligase method (Glover, White, Finnegan and Hogness, unpublished). The symbols (arrowheads, etc.) have the same meaning as in figure 2. *Drosophila melanogaster* DNA was prepared as described in figure 2 except for the omission of the actinomycin D/CsCl centrifugation. It was partially cleaved with EcoRI and fractionated by electrophoresis in a 0.7% agarose preparative slab gel at 2 volts/cm for 36 hours under conditions similar to those described by Sharp et al. (1973) for tube gels. The DNA was recovered from slices by electrophoresis, examined in the electron microscope (Davis et al., 1971), and contour lengths measured as described by Kriegstein and Hogness (1974); 40% of the linear molecules in the selected fraction (mean length and standard deviation = 14.2 ± 2.3 kb) contain EcoRI termini at both ends as judged by their ability to form circular DNA in the presence of *E. coli* DNA ligase (Mertz and Davis, 1972). This fractionated DNA was mixed with EcoRI cleaved pSC101 DNA so that the ratio of *D. melanogaster* DNA molecules with two EcoRI termini to pSC101 linear molecules was 1.3, and the joining reaction then carried out in two stages. The DNA (total concentration = $130 \mu\text{g/ml}$) was first incubated at 14°C for 45 or 110 min in 0.1 mM DPN, 1 mM EDTA, 10 mM each of $(\text{NH}_4)_2\text{SO}_4$, MgSO_4 , and Tris.HCl, pH 7.5, $14 \mu\text{g/ml}$ *E. coli* DNA ligase and $100 \mu\text{g/ml}$ BSA. This mixture was then diluted 33-fold in the reaction buffer, more DNA ligase added to a final concentration of $13 \mu\text{g/ml}$, and the incubation at 14°C continued for *c.* 16 hours. The products were fractionated as described in the text, and the fractionated DNA used in the transformation step described in figure 2, except that the recipient bacteria were *E. coli* HB101, which are *hsm*⁻, *hsr*⁻, *recA*⁻, *pro*⁻, *gal*⁻, *str*^R, F⁻ (Boyer and Roulland-Dussoix, 1969).

an equal number of pSC101 and chromosomal segments that alternate around the circle. Our observation that the mean length of the circles produced by the mixture of the two DNAs equaled the sum of their mean lengths, indicates that this number is one for most hybrids in the heterogeneous population. The following experiments indicate that this number is also one for the homogeneous pDm DNAs isolated from the transformed clones.

Three pDm DNAs (pDm1, pDm2, and pDm4) were isolated from different clones and examined to determine if they have the structures expected from the method of construction. Each exhibits a mean contour length greater than that of pSC101 DNA (Table 1), and a standard deviation that indicates they are as homogeneous with respect to length as is the parent plasmid (footnote (a), Table 1).

TABLE 1
Properties of the poly(dA):poly(dT) hybrids
(Wensink et al., 1974)

Plasmid	Contour ^a length (kb)	Heteroduplex lengths ^b		Plasmid ^c density (g/cc)	ρ_{Dm}^d (g/cc)
		Double- strand (kb)	Single- strand (kb)		
pSC101	9.22±0.06	-	-	1.7104	-
pDm1	12.09±0.11	9.3±0.2	3.1±0.1	1.7071	1.697
pDm2	17.95±0.16	9.2±0.2	8.9±0.3	1.7046	1.699
pDm4	21.84±0.19	9.3±0.2	13.0±0.2	1.7082	1.707

λ DNA (46.5 kb: Davidson and Szybalski, 1971) was used as the reference for pSC101, and pSC101 then used as reference for the pDm DNAs and for the double-stranded segment of the heteroduplexes; M13 DNA (6.6 kb: Marvin and Hohn, 1969) was used as the single-stranded reference. The standard errors given above include all measurements required to relate the sample length to the primary references, λ and M13. ^aLengths of 51 to 95 sample and reference molecules were measured from aqueous spreads. Standard deviations (s) for pSC101, pDm1, 2 and 4 are respectively, 0.34, 0.34, 0.58, and 0.62 kb; $s/m^{0.5}$ for each DNA (where m = mean length) falls in the range $0.118 \pm 0.020 \text{ kb}^{0.5}$, indicating that the DNAs are equivalently homogeneous with respect to length (Davis et al., 1971). ^bHeteroduplexes were formed and spread for electron microscopy as described in figure 4. ^cBuoyant densities were determined in CsCl as described by Champoux and Hogness (1972) and are relative to *E. coli* DNA at 1.7100 g/cc. ^d ρ_{Dm} is the buoyant density in CsCl of the Dm segment and was calculated from:

$$\rho_{Dm} = (\rho_{pDm}) / [1 + (L_{pSC} / L_{Dm})(\Delta\rho / \rho_{pSC})]$$

where ρ_{pDm} and ρ_{pSC} are the buoyant densities of the pDm and pSC101 plasmids, $\Delta\rho = \rho_{pSC} - \rho_{pDm}$, and L_{pSC} and L_{Dm} are the lengths of the pSC101 and Dm segments. L_{Dm} is taken as the difference between the pDm and pSC101 plasmid lengths. See Champoux and Hogness (1972) for the method of deriving this equation.

Heteroduplexes were formed between pSC101 and each pDm plasmid by denaturing and annealing a mixture of the two randomly nicked

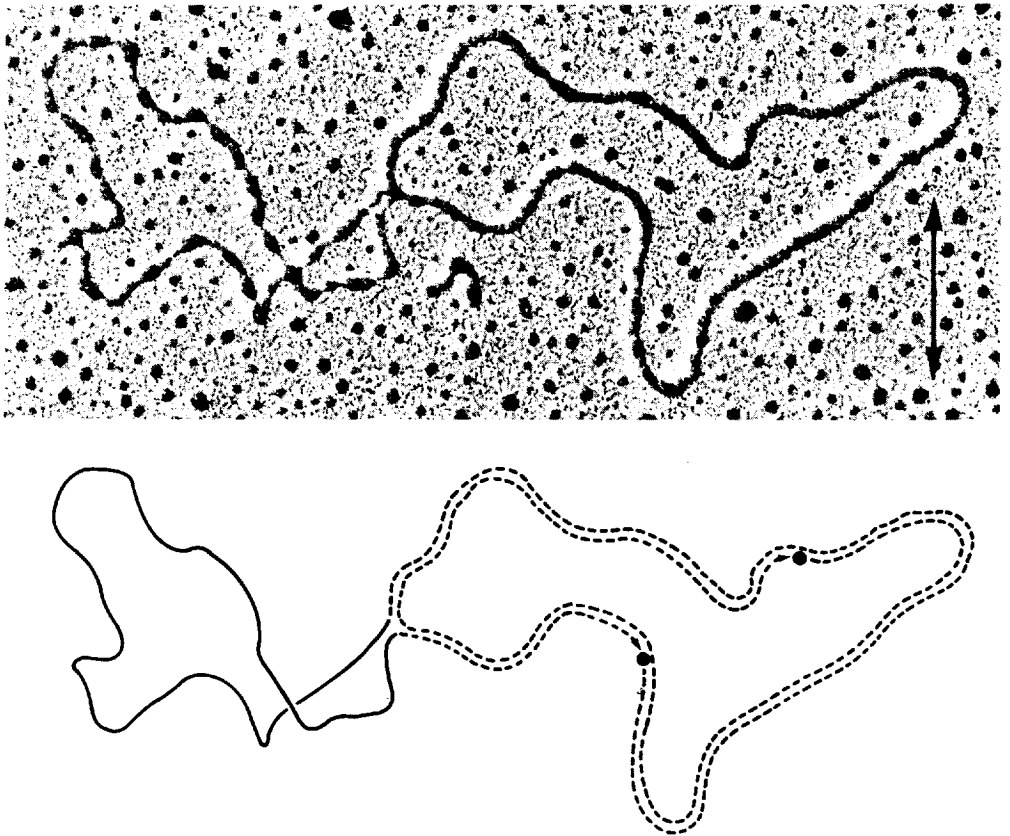


Figure 4. Heteroduplex between randomly nicked pDm2 and pSC101 DNAs (Wensink et al., 1974). The dashed and continuous lines in the diagram indicate the pSC101 and Dm segments, respectively, and the double-headed line in the micrograph represents 1.0 kb (double-stranded) or 1.2 kb (single-stranded). Denaturation and formation of the heteroduplexes in 47% formamide, and spreading in 40% formamide for electron microscopy were performed as described by Davis et al. (1971).

DNAs (Figure 4). In each case, the heteroduplexes consisted of a double-stranded ring the length of pSC101 DNA and an attached single-stranded ring, with a length equal to the difference between the lengths of the pDm and pSC101 DNAs (Table 1). Hence each pDm DNA consists of a single pSC101 segment attached to a single segment of foreign DNA, which, as will become apparent, is *D. melanogaster* DNA (i.e., a Dm segment).

If the Dm segment was inserted at the EcoRI cleavage site in pSC101 (Figure 2), then substitution of EcoRI cleaved pSC101 DNA

for the randomly nicked form in the above reaction should change the topology of the heteroduplex from two linked rings to a single ring such as that shown in Figure 5. If insertion is elsewhere,

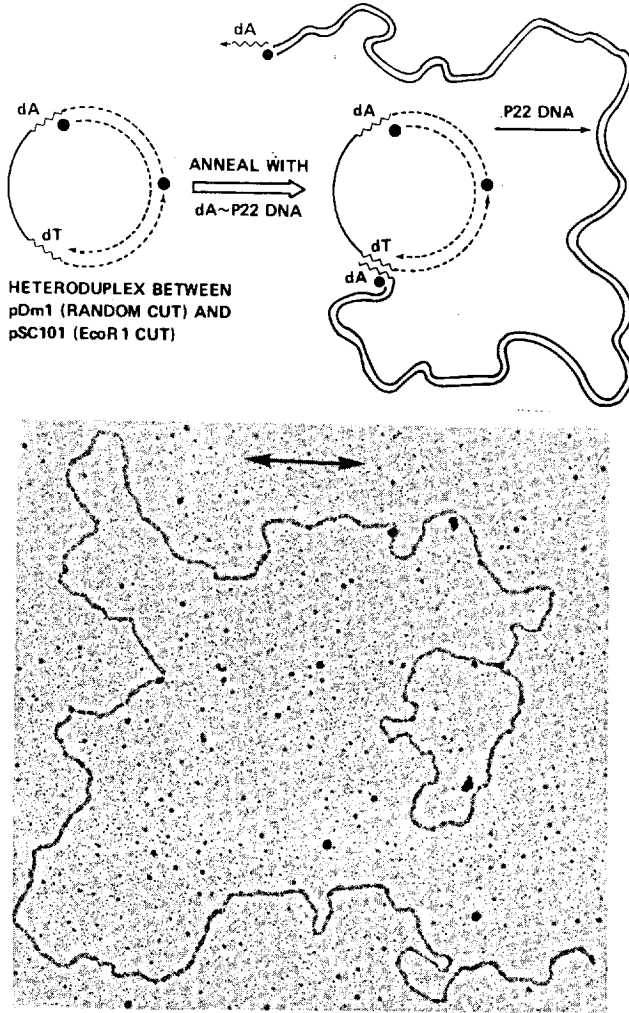


Figure 5. Attachment of poly(dA)~P22 DNA to heteroduplexes between randomly nicked pDm1 DNA and EcoRI cut pSC101 DNA (Wensink et al., 1974). Heteroduplexes were formed as indicated in figure 4, and reacted with poly(dA)~P22 DNA as described by Wensink et al. (1974). The molecule in the electron micrograph was observed after aqueous spreading (Davis et al., 1971); the single-stranded region (see diagram) is collapsed under this condition. Lengths of the circular and tail portions of such molecules are 8.9 ± 0.3 kb and 39.0 ± 0.8 kb ($N = 7$), and correspond to the lengths of pSC101 (Table 1) and P22 (Rhoades et al., 1968) DNAs. The double-headed line represents 2.0 kb of duplex DNA.

both kinds of heteroduplexes should exhibit the two-ring structure. In each case, this substitution resulted in the expected one-ring heteroduplex, consisting of single- and double-stranded segments with lengths approximately those given in Table 1 for the corresponding regions in the two-ring heteroduplexes. Insertion therefore occurred at the EcoRI site.

The results of four different experiments indicate that the two poly(dA):poly(dT) joints which effect this insertion are retained in the pDm DNAs (Wensink et al., 1974). Since space is limited, only one of these experiments will be discussed here (Figure 5). Poly(dA) extensions were added to the 3'-termini of linear phage P22 DNA by the method described in Figure 2. These molecules were then used as a probe for a single-stranded poly(dT) segment which should be present in the one-ring heteroduplexes described above if the joints are retained. Pairing between the poly(dA) probe and such a poly(dT) segment will produce a lariat which in aqueous spreads should consist of a tail the length of P22 DNA attached to a circle the length of pSC101 DNA (the single-stranded region in the circle will collapse in aqueous spreads). Such lariats were consistently observed with this probe (Figure 5), but were not observed in a control reaction where the probe was replaced by P22 DNA that had been treated with λ exonuclease but not with terminal transferase (see Figure 2). This result indicates that at least one of the expected two poly(dA):poly(dT) joints is present. The other experiments of this set confirm and extend this conclusion to include both joints.

When taken in conjunction with the known properties of the poly(dA):poly(dT) joining reaction (Jackson et al., 1972; Lobban and Kaiser, 1972), the structures of the pDm DNAs defined here provide a strong argument that their Dm segments derive from a single chromosomal segment.

Molecules constructed by the EcoRI-ligase method (Glover, White, Finnegan and Hogness, unpublished) The essential modifications of the EcoRI-ligase procedure that we have introduced are a series of molecular sizing steps which ensure that the *D. melanogaster* DNA in the hybrid derives from a single chromosomal segment. The first of these steps is the isolation of a narrow size class of chromosomal segments (mean length and standard deviation = 14.2 ± 2.3 kb; N = 126) by agarose gel electrophoresis of *D. melanogaster* DNA that was partially cleaved with EcoRI. After these molecules were allowed to react with EcoRI-cut pSC101 DNA in the presence of *E. coli* ligase (Figure 3), the reaction products were subject to a second sizing step to remove circular DNAs larger than 30 kb - i.e., to remove hybrids that contain two chromosomal segments (expected length =

37.6 \pm 3.3 kb). This was accomplished by zone sedimentation in sucrose gradients. Because the length distribution of circular DNAs in the gradient fractions was relatively broad (standard deviation = 4 to 5 kb), fractions in which the mean length was greater than 20-22 kb were excluded. Thus, the yield and purification of the 1:1 hybrids (expected length = 23.4 \pm 2.3 kb) were sacrificed in favor of eliminating the 1:2 hybrids; indeed this step favored the fractionation of the circular pSC101 dimers which have a length of 18.4 kb.

The DNA in the combined gradient fractions was then used to transform tetracycline sensitive *E. coli* to resistance. The hybrid plasmids were first isolated from a mass culture that was inoculated with the bacteria from *c.* 800 transformed colonies, and these hybrids then used to obtain individual clones in a second transformation step. This procedure was adopted because the ratio of circular hybrids to circular pSC101 monomers and dimers in the gradient fractions was estimated from the length distributions to be quite small (*c.* 0.1), and because a trial screening indicated that only 2-3% of the 800 transformants contained hybrid-size plasmids. The plasmid DNAs from the mass culture fell into two size classes, corresponding to pSC101 monomers and to hybrid plasmids. As expected, the hybrid class constituted only a few percent of the plasmids, and had a mean length and standard deviation of 22.6 \pm 2.7 kb (N = 37). This DNA was separated from the pSC101 monomers by electrophoresis and then used to obtain individual clones of transformed cells.

Seven of these clones were selected and their plasmid DNAs characterized with respect to the properties listed in Table 2. The contour lengths of each of these pDm DNAs is within the range expected for the 1:1 hybrids, but is outside the range expected for the 1:2 hybrids. Only six pDm plasmids are listed in Table 2 because two (pDm101 and 104) exhibit the same properties. We presume that they both originate from the same one of *c.* 20 hybrid clones in the original population of 800 transformants.

Following complete cleavage with EcoRI, each pDm DNA yielded a fragment of pSC101 size together with the fragments shown in Table 1, which correspond in total length to the presumptive Dm segment. Whereas the pSC101 size fragment was obtained in equimolar amounts with respect to the other fragments from five of the pDm DNAs, pDm 101 yielded 2 moles of this fragment per mole of the others. We have assigned one of these pSC101 size fragments in pDm101 to the Dm segment (the 9.2 kb fragment in Table 2) because: (1) two pSC101 segments would occupy 18.4 of the 24.6 kb in the hybrid and leave only 6.2 kb for the chromosomal segment, which is more than three

standard deviations less than the mean length of the chromosomal segments used to construct the hybrid; and (2) the reassociation kinetics of pDm101 DNA in the presence of *D. melanogaster* DNA (next section) indicate that *c.* 60% of the pDm101 DNA consists of *D. melanogaster* sequences (the expected amounts are 63 and 25% for a 1:1 and 2:1 hybrid, respectively). We conclude that each hybrid contains a single pSC101 segment.

TABLE 2

Properties of the EcoRI-ligase hybrids (Glover, White, Finnegan and Hogness, unpublished)

Plasmid	L_{pDm}^a (kb)	L_{Dm}^b (kb)	EcoRI fragments from Dm seg- ment ^c (kb)	Plasmid density ^d (g/cc)	ρ_{Dm}^d (g/cc)
pDm101	24.6±0.2	15.4	<i>9.2, 3.8</i> <i>1.5, 1.4</i> $\Sigma = 15.9$	1.7045	1.701
pDm102	27.7±0.3	18.5	11.8, 6.1 $\Sigma = 17.9$	1.7065	1.705
pDm103	26.3±0.3	17.1	17.0	1.7047	1.702
pDm105	26.8±0.3	17.6	4.9, 3.3 <i>2.9, 2.9</i> <i>2.6, 1.4</i> $\Sigma = 18.0$	1.7029	1.699
pDm106	24.5±0.4	15.3	7.9, 6.1 1.3 $\Sigma = 15.3$	1.7069	1.706
pDm107	27.2±0.2	18.0	8.6, 4.9 <i>4.3</i> $\Sigma = 17.8$	1.7058	1.703

^a L_{pDm} is the contour length of the plasmid and was determined as described in Table 1, except that 17 ± 4 molecules were measured.

^b $L_{Dm} = L_{pDm} - L_{pSC}$, where L_{pSC} is the contour length of the pSC101 plasmid taken from Table 1.

^cThe lengths of the fragments are the electron microscopic contour lengths except for the italicized values, which were determined by agarose gel electrophoresis using the EcoRI fragments of phage λ DNA (Thomas and Davis, 1975) and the EcoRI and HpaI fragments of SV40 DNA (Berg et al., this volume) as length standards.

^dThe values were obtained as described in Table 1.

The EcoRI fragments from the Dm segment indicate that it consists of only one chromosomal segment. This is clearly the case for pDm103 where the Dm segment contains no internal EcoRI sites. Were any of the other Dm segments to derive from two chromosomal segments, then one of the two must have a length less than or equal to 6.7, 6.1, 8.9, 7.4 and 8.6 kb for pDm101, 102, 105, 106 and 107 respectively. Since these values are more than two, and in some cases more than three standard deviations less than the mean length of the chromosomal segments used for hybrid construction, it is highly improbable that segments of this size were inserted in these pDm DNAs.

We conclude that all pDm DNAs given in Tables 1 and 2 consist of a single chromosomal segment linked to a single pSC101 DNA. Although the EcoRI-ligase joining method appears simpler at first sight, we find that the sizing steps required to ensure single segment insertion make it more complex and less efficient than the poly(dA):poly(dT) method. These sizing steps could be eliminated if one were willing to limit sequence mapping to chromosomal segments that do not contain internal EcoRI sites. However, this would restrict most maps to small segments representing only a fraction of the average chromomeric unit.

Identification and mapping of the sequences in the Dm segment

Tables 1 and 2 give the buoyant densities of the plasmid DNAs in neutral CsCl gradients (experiments on pDm1, 2 and 4 are from Wensink et al., 1974; those on pDm101-107 are from Glover, White, Finnegan and Hogness, unpublished). The density of the DNA in each Dm segment has been calculated from these measured densities and the known lengths of the Dm and pSC101 segments, with results that are also given in these tables. In each case, the Dm density falls within the density profile of the main band of *D. melanogaster* DNA (Schachat and Hogness, 1974). More critical evidence that the Dm segments consist of *D. melanogaster* sequences is given by the following experiments which also identify these sequences with respect to their frequency of repetition and their location in the *D. melanogaster* genome.

Self-driven reassociation kinetics of the pDm DNAs The reassociation kinetics of each pDm DNA has been determined to test for sequence repetition within the Dm segment. Figure 6 shows these kinetics for the pSC101, pDm1, 2 and 4 DNAs. The data points fit the modified second order rate equation,

$$C/C_0 = (1 + kC_0 t)^{-n} \quad (1)$$

(sigmoid curve in Figure 6), or its equivalent,

$$(C_0/C)^{1/n} = 1 + kC_0 t \quad (2)$$

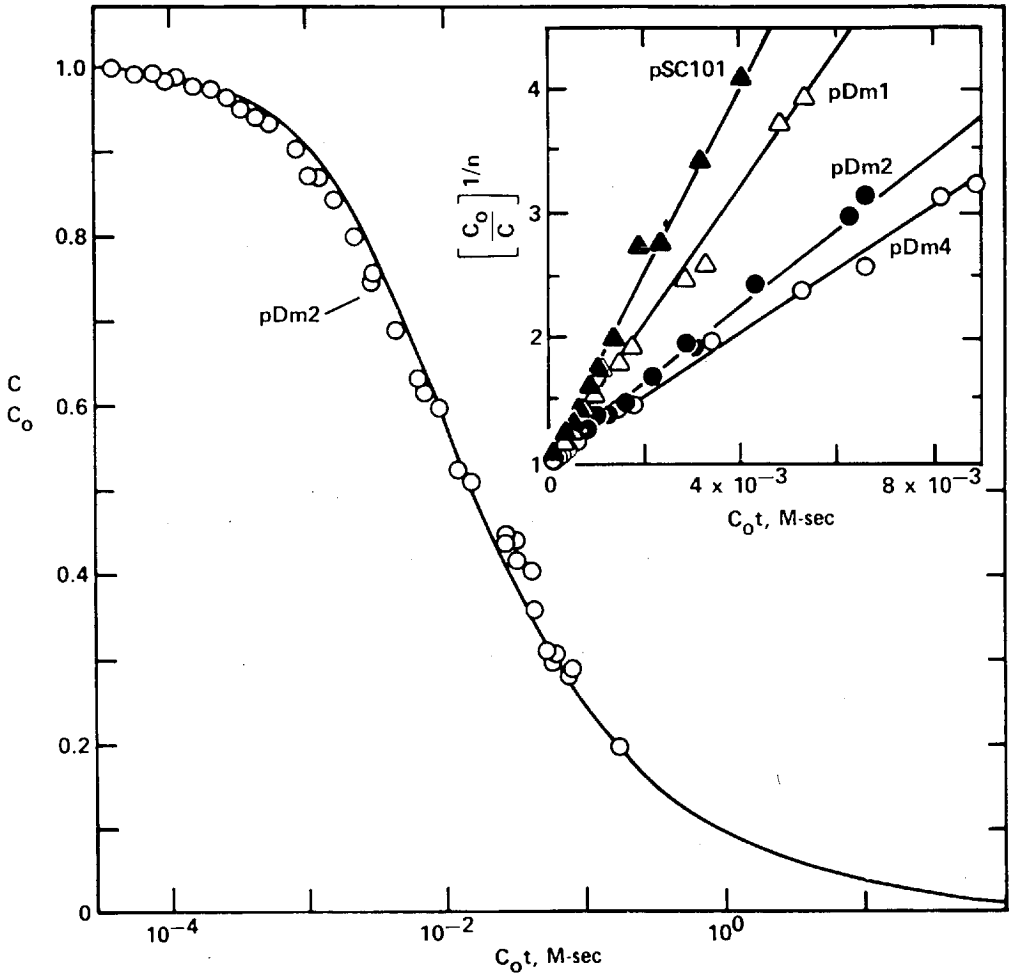


Figure 6. Self-driven reassociation kinetics of plasmid DNAs (Wensink et al., 1974). Plasmid DNAs were labelled *in vitro* with ^{32}P to $1 - 2 \times 10^8$ cpm/ μg (Schachat and Hogness, 1974) and mixed with a 100-fold excess of unlabelled plasmid DNA to obtain a C_0 of $c. 2 \times 10^{-7}$ M in the reassociation reaction. Shearing, denaturation and reassociation (assayed with S_1 endonuclease) were carried out in the presence of calf thymus DNA (250 $\mu\text{g}/\text{ml}$) as described by Schachat and Hogness (1974). Reassociation occurred at 65°C in 0.5 M NaCl in TE. The sigmoid curve and the straight lines in the inset are the least squares best fit of equations (1) and (2) to the data. The second order rate constants for the plasmids shown here and for the six pDm DNAs listed in Table 2 are: $k_{\text{pSC}} = 539$; $k_{\text{pDm1}} = 415$; $k_{\text{pDm2}} = 264$; $k_{\text{pDm4}} = 189$; $k_{\text{pDm101}} = 241$; $k_{\text{pDm102}} = 258$; $k_{\text{pDm103}} = 207$ (slow component); $k_{\text{pDm105}} = 308$; $k_{\text{pDm106}} = 339$; $k_{\text{pDm107}} = 163 \text{ M}^{-1} \text{ sec}^{-1}$.

(inset to Figure 6), where C and C_0 are the concentrations of single-stranded DNA at time t , and $t = 0$; k is the second order rate constant; and n is an empirically determined constant that takes account of secondary reactions detected when the reassociation of random fragments is assayed with the single-strand specific S_1 endonuclease ($n = 0.44$; Morrow, 1974).

The reassociation of nonrepetitive DNAs has been observed to fit equation (1) when assayed by S_1 , and to yield values of k that are inversely proportional to the size of the genome from which the DNA was obtained (Morrow, 1974). Given that pSC101 is nonrepetitive, then we expect that,

$$k_{pDm} = k_{pSC} (L_{pSC}/L_{pDm}), \quad (3)$$

if the sequences in the Dm segment are also nonrepetitive (the subscripts in equation (3) designate the plasmid DNA, and L is the length of that DNA). If, on the other hand, the Dm sequences are repeated within the segment, then the observed k_{pDm} will be greater than that predicted by equation (3). Indeed, if these sequences are repeated many times, a biphasic curve should be generated due to the more rapid reassociation of these repeated sequences.

The reassociation kinetics of pDm101, 102, 105, 106 and 107 are like that observed for pDm1, 2 and 4 (Figure 6) in that the data points fit equation (1) within experimental error. The ratio of the observed k_{pDm} to that expected for a nonrepetitive Dm segment [i.e., that calculated from equation (3)] is 1.01, 0.95, 0.84, 1.19 and 0.89 for pDm1, 2, 4, 101 and 107, respectively. These values are not significantly different from one, and indicate that there is little if any sequence repetition within the Dm segments of these five plasmids.

The ratios for pDm102, 105 and 106 are 1.44, 1.66 and 1.67, respectively. Since it is unlikely that these deviations to the high side of one are due to experimental variation, they suggest that some sequence repetition may occur in these Dm segments, although the frequency of that repetition cannot be large. For example, we have calculated that if all sequences in these Dm segments are repeated twice per segment, the k_{pDm} will be increased 1.5- to 1.6-fold above that expected for no repetition; and a triple repetition of all sequences will increase the k_{pDm} about 2.0-fold. In neither case does the computed reassociation curve exhibit a sufficient biphasic character that it could be distinguished experimentally from the curve generated by equation (1).

The distribution of the EcoRI fragments derived from these Dm segments also places some restrictions on sequence repetition within them. Table 2 shows that among all the length classes of EcoRI

fragments obtained from Dm102, 105 or 106, only one (the 2.9 kb fragments from Dm105) contains more than one member. If we assume that the postulated sequence repetition in these hybrids results from tandemly repeated sequences which occupy most of the Dm segment (i.e., more than two-thirds), then we can deduce from the distribution of the EcoRI fragments that the number of repeating unit per hybrid must be ≤ 2 in each case. The size of the repeating unit in the *D. melanogaster* genome would then be ≥ 6 to 9 kb. Thus the kinetics and the EcoRI fragments are compatible with a 2-fold repetition that occupies most of the Dm segment, but are not compatible with a 3-fold or greater repetition of this extent. We defer a consideration of other possible forms of sequence repetition until the last section of this paper.

The self-driven reassociation curve for the remaining hybrid, pDm103, exhibits a biphasic character that indicates a small fraction of the sequences in this hybrid renature much faster than the remainder. Using a procedure that is essentially the same as that described in Figure 8, we find that: (1) about 10% of the pDm103 sequences renature at such a fast rate that they must be repeated on the order of 100 times per molecule; and (2) the second order rate constant for the remaining 90% equals that expected for non-repetitive sequences (i.e., ratio of observed to expected constant 0.99). Hence, about 15% of the Dm103 segment, or 2.6 kb, appears to consist of a large number of small repeating units (c. 30 base pairs per unit), and c. 85% or 14.5 kb, consists of sequences which are not repeated within the hybrid. We are particularly interested in these inferences because the Dm103 segment contains one representative of the gene for the 18 and 28S rRNAs, as will be demonstrated in the following sections.

Drosophila DNA driven reassociation kinetics of the pDm DNAs The reassociation kinetics of a minute amount of ^{32}P -labelled pDm DNA in the presence of a very large amount of *D. melanogaster* DNA (w/w 10^6 to 10^7) have been examined to determine the fraction of the pD sequences that derive from *D. melanogaster*, and to measure the repetition frequency of these sequences in the *D. melanogaster* genome. These experiments are illustrated by the examples given in Figure

The reassociation rate of pSC101 DNA is not affected by the *D. melanogaster* DNA, as is seen in Figure 7A by the agreement between the data points and the curve computed from equation (1), using the value of k_{pSC} determined in Figure 6 (the reassociation of pSC101 DNA appears to occur at high C_0t in Figure 7A because the C_0 in the abscissa refers to the *D. melanogaster* DNA). By contrast, the reassociation rate of a fraction of each pDm DNA is increased by the addition of the *D. melanogaster* DNA; and this fraction corresponds

to the fraction of the hybrid that is occupied by the Dm segment. This effect is illustrated in Figure 7A by the biphasic curve obtained with the pDm1 plasmid, where 24% of the DNA is represented by the Dm segment. One-fourth of the labelled sequences can be observed to reassociate rapidly along with the moderately repetitive sequences of *D. melanogaster* DNA, while the remaining three-fourths exhibit the kinetics expected for pSC101 DNA.

Given (i) the concentration of labelled pDm DNA, (ii) the fraction of this DNA that consists of the pSC101 sequences (Tables 1 and 2), and (iii) the value of k_{pSC} , then equation (1) can be used to compute the contribution of the pSC101 sequences to the total reassociated DNA for each datum point. The values of C^*/C_0^* for the total labelled DNA (ordinate to Figure 7A) can then be transformed to C^{Dm}/C_0^{Dm} , where C^{Dm} represents the single-stranded concentration of labelled *D. melanogaster* sequences from the plasmid. A computer program was written to effect this transformation for each data point and the results are given in Figure 7B for pDm1, 2 and 4, along with curves that represent that best fit of equation (1) to the data.

The good fit between the curves and the transformed data points suggests that the sequences within the Dm segments of these three plasmids form a homogeneous population with respect to their repetition frequency, α , within the *D. melanogaster* genome. It is convenient to classify such Dm segments as belonging to the 1α class. When the same experiment was performed with the other six pDm DNAs, the same good fit was observed for four (pDm102, 105, 106, 107), which are therefore also placed in the 1α class. The value of α for each Dm segment in this class can be calculated from the following equation (Wensink et al., 1974):

$$\alpha = (k'_{pDm}/k_{pDm})(L_D/L_{pDm}) \quad (4)$$

where k_{pDm} and L_{pDm} have their previously assigned meanings; k'_{pDm} is the value of k obtained when equation (1) is fitted to the data in Figure 7B; and L_D is the length of DNA in the *D. melanogaster* genome (165,000 kb).

The resulting values of α for pDm1, 2, 4, 102, 105, 106 and 107 are given in Table 3. Equation (4) is valid only if the Dm sequences are not repeated within the Dm segment. If it is assumed that the Dm segments consist of tandemly repeated subunits, then the value of α for these repeated units is given by equation (5), which is derived in the legend to Table 3;

$$\alpha = (k'_{pDm}/k_{pSC})(L_D/L_{pSC}) \quad (5)$$

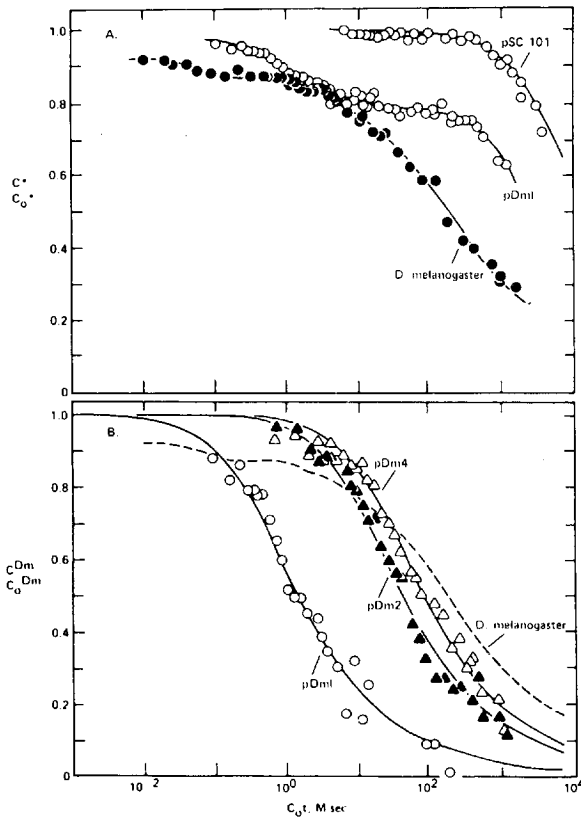


Figure 7. Reassociation kinetics of plasmid DNAs in the presence of *D. melanogaster* DNA (Wensink et al., 1974). Reassociation was carried out as described in Figure 6 except that the unlabelled pDm DNA and calf thymus DNA were omitted and *D. melanogaster* DNA was added to a concentration ($c. 1 \times 10^{-2} M$) that was 7×10^5 times that of the labelled plasmid DNA. [The ratio used for the pDm101 - 107 plasmids (not shown in the figure) was $c. 10$ -fold greater; Glover, White, Finnegan and Hogness, unpublished]. In the abscissa C_0 refers to the *D. melanogaster* DNA. (A) The ordinate, C^*/C_0^* , refers to the labelled plasmid DNA for the pSC101 and pDm1 curves, and to the labelled *D. melanogaster* DNA for the third curve (plasmid DNA absent). (B) The ordinate C^{Dm}/C_0^{Dm} , represents the labelled Dm sequences in the pDm DNA (see text). The curves are the least squares best fit of equation (1) to the data points in this figure (i.e., $C^{Dm}/C_0^{Dm} = (1 + k'_{pDm} C_0 t)^{-n}$, where C_0 refers to the added *D. melanogaster* DNA), and correspond to: $k'_{pDm1} = 2.74$; $k'_{pDm2} = 0.101$, and $k'_{pDm4} = 0.049 M^{-1} sec^{-1}$. The data points for pDm1 are from one of the two experiments used to provide the data for the pDm1 curve in panel A. The values of k'_{pDm} for pDm102, 103, 105, 106 and 107 are 0.039, 16.1, 0.082, 0.100 and 0.036 $M^{-1} sec^{-1}$.

This equation is valid for any number of tandem repeats that encompass the Dm segment, and is equivalent to equation (4) when equation (3) can be applied, i.e., when there is no sequence

repetition in the Dm segment. The values of α calculated from equation (5) are given in parentheses in Table 3 for pDm102, 105 and 106, i.e., for those cases where the self-driven reassociation kinetics suggest that there may be a tandem repetition within the Dm segment. Although equation (4) is less general than (5), it has the advantage that both rate constants in the ratio, $k'_{\text{pDm}}/k_{\text{pDm}}$, are determined with the same preparation of labelled pDm DNA. This leads to a cancelling out of the effects of such factors as the length distribution of the reassociating fragments, and their base composition and sequence. This is not the case for equation (5), which therefore allows a greater potential error in evaluating α .

The reassociation kinetics of pDm101 DNA in the presence of *D. melanogaster* DNA yields a biphasic curve (Figure 8) which can be most simply interpreted by dividing the Dm101 sequences into two populations of approximately equal size which exhibit α values that differ by about two orders of magnitude (Table 3). Following a procedure similar to that described by Morrow (1974) (see legend to Figure 8), we find that $\alpha = 30$ for about 43% of the Dm101 sequences, and that $\alpha = 0.2$ for the remainder. The low value of α observed for the slow component in Figure 8 deserves comment. Since the *D. melanogaster* DNA that was used to drive the reassociation reaction is half derived from males and half from females, the value of α for a unique sequence in the Y and X chromosomes is expected to equal 0.25 and 0.75 respectively, whereas α should equal 1.00 if the sequence is located in an autosome. Additional arguments supporting this suggestion that the Dm101 segment derives from the Y chromosome are discussed in the last section of this paper.

The quantitative analysis of the reassociation of pDm103 DNA in the presence of *D. melanogaster* DNA has proved difficult due to an exceptional scatter of the data points that is reproduced in repeated experiments. This scatter has prevented an assignment of the Dm103 segment to the 1α class or to the 2α class occupied by Dm101, although the fast and slow components observed in the self-driven reassociation reaction suggest that it may belong to the 2α class. Since the data points scatter about the best fit curve of equation (1) in a reasonable manner (i.e., the shape of the scatter envelope corresponds to this curve) we have used the single k'_{pDm} provided by this curve and equation (4) to calculate a repetition frequency of 400 for the Dm103 sequence (Table 3).

Mapping the Dm sequences by *in situ* hybridization to polytene chromosomes ^3H -labelled RNA copies (cRNA) of the sequences in each pDm DNA and in pSC101 DNA were transcribed with *E. coli* RNA polymerase and hybridized *in situ* to *D. melanogaster* polytene chromosomes. Figure 9 illustrates the results of this procedure for pDm4 and Table

TABLE 3

Repetition frequency, α , and chromosomal location of the Dm sequences (Wensink et al., 1974; Glover, White, Finnegan and Hogness, unpublished)

Dm segment	Length of segment (kb)	α^a	Location ^b
<u>Group I:</u>			
Dm102	18.5	0.9 (1.3)	X: 8A
Dm107	18.0	1.3	2R: 41D
Dm4	12.6	2.0	3L: 62E
Dm105	17.6	1.6 (2.7)	3R: 82F
Dm106	15.3	2.0 (3.3)	2R: 41D
Dm2	8.7	3.5	3R: 84D
<u>Group II:</u>			
Dm101	15.4		Chromocenter
43%	6.6	30	
57%	8.8	0.2	
Dm1	2.9	90	Chromocenter (β -hetero-chromatin) + 15 chromomeric regions: X: 2A, 18C, 19E 2L: 27E, 32B 2R: 43D, 56F, 59B, 59F 3L: 63A 3R: 84A, 87E, 88B, 97B, 99D-E
Dm103	17.1	400	Nucleolus

^a α was calculated from equation (4) except for those values in parentheses which were calculated from equation (5) which is derived as follows. The definition of α for a tandemly repeating subunit is given by $\alpha = FL_D / (fL_{pDm} / r)$, or $\alpha = r(F/f)(L_D / L_{pDm})$, where r is the number of subunits in Dm, F and f are, respectively, the fraction of L_D and L_{pDm} represented by these subunits. The second order rate constant for the repeating subunit, k_r , is given by $k_r = r[(1-f)/f]k_{pSC}$, or by $k_r = k'_{pDm} / (F + fR)$, where R is the ratio of pDm to *D. melanogaster*

3 lists the results obtained for all nine pDm hybrids. As expected, the pSC101 cRNA exhibited no *in situ* hybridization.

Table 3 shows that the pattern of *in situ* hybridization is correlated with the repetition frequency. The hybrids are listed in order of increasing α values, and broken into two groups where there is an order of magnitude jump in this value. The first group of six exhibit α values close to one (≤ 3.5), and display a simple pattern in which the silver grains are restricted to a single chromomeric region. None of the second group with the higher α values display this simple chromomeric pattern, and indeed, each of its three members contain sequences found in the heterochromatic portion of the *D. melanogaster* chromosomes. These groups are discussed in the next and final section of this paper.

CONCLUDING ARGUMENTS

Dm segments in group I

Two-thirds of the hybrids isolated by either of the two procedures fall into group I (Table 3), and the six members of this group account for all but one of the Dm segments that contain sequences located in the chromomeric units. These frequencies suggest that the majority of chromomeric segments in *D. melanogaster* exhibit the characteristics that define group I - i.e., most, if not all sequences in the segment are restricted to one or a few adjacent chromomeric units, and are either not repeated or are repeated only a few times within this chromomeric region.

The apparent high frequency and the characteristics of group I segments provide an argument against explanations of the chromomere anomaly that are based on tandem repetition (Callan, 1967; Thomas, 1970; Lee and Thomas, 1973). First consider the three Dm segments in group I that do not exhibit internal sequence repetition as judged by their self-driven reassociation kinetics. These are Dm 2, 4 and 107 and their respective α values are 3.5, 2.0 and 1.3 repeats per genome. Since we are not confident that these α values are accurate to better than a factor of two, each of these segments could be part of a tandem array of a small number of repeating units. However, if this is the case, then the repeating unit is at least as large as the segment, or 8.7, 12.6 and 18.0 kb, respectively. That

DNA used in the determination of k'_{pDm} (e.g., in figure 7B). Hence, $(F + fR)/f = (1/r)[1/(1/f)(k'_{pDm}/k_{pSC})]$. Since $F \gg fR$ and $1/(1-f) = L_{pDm}/L_{pSC}$, then $F/f = (1/r)(L_{pDm}/L_{pSC})(k'_{pDm}/k_{pSC})$, which can be substituted in the equation defining α to obtain equation (5).

^bSee Lindsley and Grell (1968) for the numeration of these chromomeric regions which contain from 4 to 17 bands. The autoradiographic resolution is to within a few bands, and hence in many cases better than to one of the indicated regions.

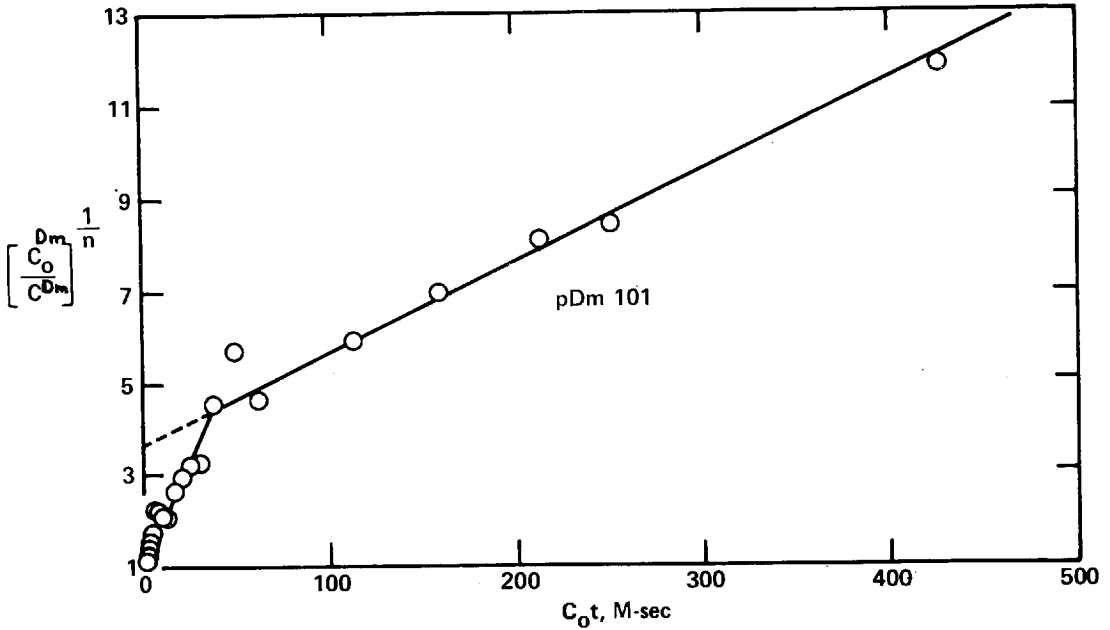


Figure 8. Biphasic reassociation kinetics of pDm101 in the presence of *D. melanogaster* DNA. The reassociation reaction was carried out as described in Figure 7 and the contribution of the DNA in the pSC101 segment subtracted from each data point to yield the ordinate values of $(C_O^{Dm}/C^{Dm})^{1/n}$, where C^{Dm} and C_O^{Dm} represent the concentration of the labelled single-stranded Dm DNA at time, t , and $t = 0$ (see text and Figure 7). The data are plotted according to equation (2) with $n = 0.44$ rather than equation (1) to facilitate the following analysis. Let $C_O^{Dm} = C_O^1 + C_O^2$ and $C^{Dm} = C^1 + C^2$, where the 1 and 2 superscripts refer, respectively, to the fast and slow renaturing DNA in the labelled Dm segment. In the presence of a high concentration of *D. melanogaster* DNA, $C^1/C_O^1 = (1 + k_1 F_1 C_0 t)^{-n}$ and $C^2/C_O^2 = (1 + k_2 F_2 C_0 t)^{-n}$, where C_0 is the concentration of the added *D. melanogaster* DNA, F_1 and F_2 are the fractions of that DNA represented by the fast and slow sequences, and k_1 and k_2 are the second order rate constants of the fast and slow components. These equations follow from equation (1) if $F_1 C_0 \gg C_O^1$ and $F_2 C_0 \gg C_O^2$, which is the case. Hence,

$$C_O^{Dm}/C^{Dm} = C_O^{Dm}/[C_O^1(1 + k_1 F_1 C_0 t)^{-n} + C_O^2(1 + k_2 F_2 C_0 t)^{-n}] \quad \dots (6)$$

When $C_0 t$ is large, $C^1 = C_O^1(1 + k_1 F_1 C_0 t)^{-n} \rightarrow 0$, and equation (6) becomes $(C_O^{Dm}/C^{Dm})^{1/n} = (C_O^{Dm}/C_O^2)^{1/n} (1 + k_2 F_2 C_0 t)^{-n}$; i.e., the data points for large $C_0 t$ ($C_0 t > 100$ M. sec) should fit a straight line with intercept, $1 = (C_O^{Dm}/C_O^2)^{1/n}$, and slope, $S = k_2 F_2 (C_O^{Dm}/C_O^2)^{1/n}$. Hence $1 - I^{-n} = C_O^2/C_O^{Dm}$ and $1 - I^{-n} = C_O^1/C_O^{Dm}$, which are, respectively, the fractions of the Dm segment represented by the slow and fast renaturing sequences. The repetition frequency of the slow Dm DNA in the *D. melanogaster* genome, α_2 , is given by $\alpha_2 = (F_2/f_2)(L_D/L_{pDm})$, where f_2 is the fraction of pDm DNA represented by the slow sequences. Since $k_2 = k_{pDm}/f_2$, we have $\alpha_2 = (k_2 F_2/k_{pDm})(L_D/L_{pDm})$, which can be used to calculate α_2 , since $k_2 F_2 = S/I$. Similarly, $\alpha_1 = (k_1 F_1/k_{pDm})(L_D/L_{pDm})$, where α_1 is the repetition frequency of the fast Dm DNA in the *D. melanogaster* genome. Given C_O^1/C_O^{Dm} , C_O^2/C_O^{Dm} and $k_2 F_2$, then $k_1 F_1$, and hence α_1 , can be determined by solving equation (6) for data points at low $C_0 t$ values; i.e., < 10 M. sec.

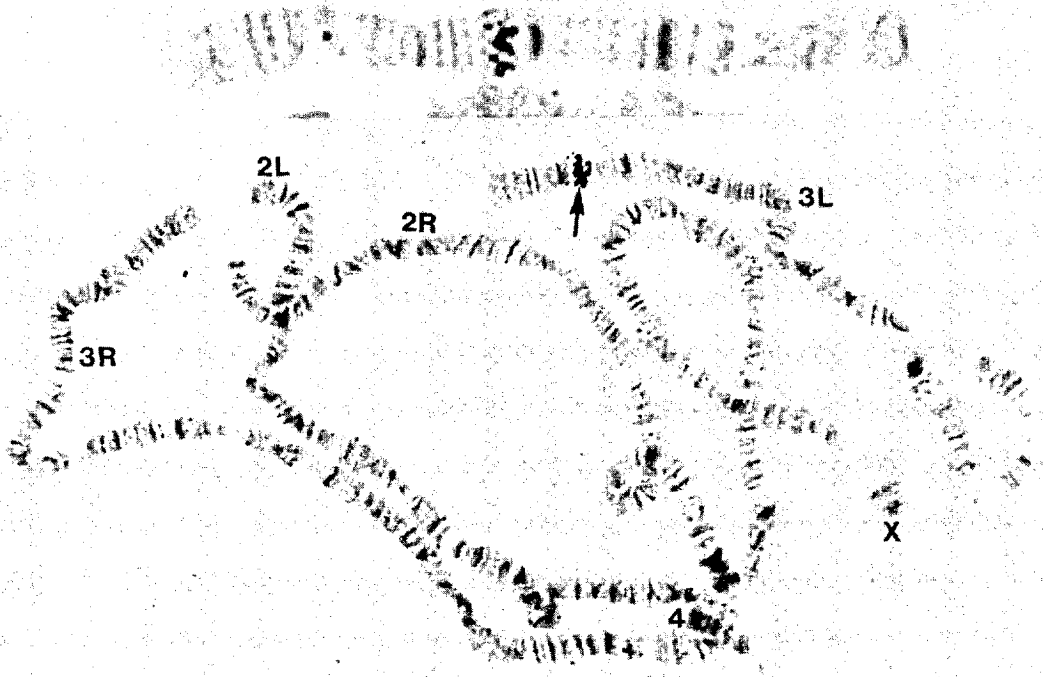


Figure 9. *In situ* hybridization of pDm4 cRNA (Wensink, et al., 1974). ^3H -cRNA was prepared by transcription of pDm4 DNA with *E. coli* RNA polymerase, and hybridized to salivary gland chromosomes prepared from *gt/gt*^{X11} larvae as described by Wensink et al. (1974). Exposure times were 13 and 41 days for the upper (2200x) and lower (550x) micrographs, respectively. The silver grains are clustered over the 62E region in 3L, i.e., the left arm of chromosome 3 (see Lindsley and Grell, 1968). This arm has broken away from the chromocenter, which is located at the lower right of the figure.

is, the repeating unit would have dimensions comparable to a chromomeric unit itself, rather than to the smaller coding sequences which, according to tandem repetition models are repeated many times within a chromomeric unit.

The same argument applies to the other three segments in group I (Dm102, 105 and 106) even though their self-driven reassociation kinetics suggest some internal sequence repetition. This is because we were able to show that a tandem array which occupied most of one of these segments could not contain more than two repeating units within the segment, a conclusion that is also supported by the low α values given in Table 3, whether calculated by equation (4) or

(5). Thus, although a tandemly repeating unit that accounts for these α values could have a length less than that of the segment, it must still be greater than 6 to 9 kb.

We therefore arrive at the general conclusion that if the group I segments derive from tandem arrays, then the repeating unit in such an array must be ≥ 6 to 18 kb, the value depending upon the particular segment. The molecular weight of a polypeptide coded by such a unit would be greater than 220,000 to 660,000 daltons, and although structural genes of this size are not impossible, it is highly unlikely that all six segments in this group would derive from such large genes.

How can one explain the data for such cases as Dm2, where the evidence for repetition of the segment is reasonably convincing (i.e., where the α value is significantly different from one)? Perhaps Dm2 represents a duplication (or triplication, etc.) of an entire chromomeric unit. Lefevre and Green (1972) have described a genetic duplication in the 3C region that appears to be of this kind, and Lefevre (1974) has noted other regions where such duplications probably occur, each containing a pair of adjacent dark-staining doublet bands that characterize the duplication in 3C. We note that the Dm2 segment is located in a region containing such a pair of doublets (84D 1-2 and 3-4) and that the silver grains are centered over this part of the 84D region.

The last point about group I segments that we wish to mention is that the *Drosophila* DNA driven reassociation kinetics (Figure 7B) of each indicates the absence of a short period interspersion of repetitive sequences of the dimensions observed for *Xenopus laevis* (Davidson et al., 1973) and sea urchins (Graham et al., 1974). The average interspersion pattern in these two animals consists of 0.3 kb stretches of repetitive sequences that alternate with 0.7 to 1.0 kb stretches of nonrepetitive sequences. Since these repetitive sequences constitute about one-fourth of this interspersion pattern and exhibit repetition frequencies more than an order of magnitude greater than the α values for group I segments, they should cause the data points at lower C_0t values to fall well below the fitted curves for equation (1) (Figure 7B).

Although such a marked fall off of the data points was not observed for any segment in group I, it should be emphasized that an interspersion pattern in which the stretch of repetitive sequences is considerably shorter, or in which the interspersion period is considerably longer, would not be reliably detected by these reassociation kinetics. For example, the fall off of the early data points from the pDm4 curve (Figure 7B) would indicate a more rapidly re-

associating minority fraction representing about one-tenth of the segment were it not for the fact that the deviation of these points from the curve is at the borderline of significance.

Dm segments in group II

Dm101 The observation that Dm101 contains both nonrepetitive and repetitive sequences is compatible with the interspersion of repetitive sequences of the kind discussed above. However, each of these two kinds of segments could also be localized in single blocks within the segment. These different patterns can be distinguished by an examination of the reassociation kinetics of each EcoRI fragment (Table 2) in the presence of *D. melanogaster* DNA and by ordering these fragments within the segment (e.g., by heteroduplex mapping). We have not as yet carried out this kind of internal sequence mapping of particular segments, and this example is mentioned here only to indicate the general concept.

The sequences in the Dm101 are restricted to the chromocentric heterochromatin, as judged by the results of *in situ* hybridization (Table 3). This distribution is compatible with the previous suggestion that the Dm101 segment derives from the Y chromosome, since the heterochromatic Y is restricted to the chromocenter of the polytene cluster (Rudkin, 1972). Although the sequences in Dm101 with $\alpha = 0.2$ would, according to this suggestion, necessarily be restricted to Y, the sequences with $\alpha = 30$, which represent c. 200 kb in the genome, could be located both in Y and in the centromeric heterochromatin of other chromosomes, since this heterochromatin is also restricted to the chromocenter.

Dm1 The sequences in Dm1, which are repeated about 90 times and represent about 260 kb in the genome, are found in the chromocentric heterochromatin and in at least 15 different chromomeric regions (Table 3); i.e., these sequences are clearly interspersed at very long intervals within the chromosomal DNA.

There are two extreme interpretations of these results. All sequences within the Dm1 segment could be located in each of the 16 chromosomal regions. Or these sequences could be divided into 15 subsets such that one chromosomal region contains all 15 (i.e., the chromosomal segment isolated in pDm1), and the other regions each contain a different one of the subsets. The simplest method for distinguishing among these interpretations is to divide the Dm1 segment into subsegments, and to determine for each isolated subsegment the chromosomal location of its sequences by *in situ* hybridization.

An analysis of the function of the Dm1 sequence must await this kind of sequence mapping. At this stage we wish to point out

only that the wide but specific distribution of these sequences in the *D. melanogaster* chromosomes can be fitted to many regulatory models for the eukaryotic cistron of which the prototype is that proposed by Britten and Davidson (1969). In this regard, we note that the pattern of silver grains over the 56F region, that is given by the Dml sequences (Table 3), is indistinguishable from that given by ³H-labelled 5S ribosomal RNA [M.L. Pardue, personal communication; also see Wimber and Steffensen (1970) and Grigliatti et al. (1974)]. Perhaps certain of the sequences in Dml are involved in the regulation of the expression of the 5S genes.

Dml03 The high repetition frequency of Dml03 and the localization of its sequences within the nucleolus provided the initial suggestion that this segment might contain the gene for the 18-28S rRNA. Proof that this is the case has recently been provided by the hybridization of labelled 18 and 28S rRNA from *D. melanogaster* cell cultures to nitrocellulose filters containing single strands of pDml03 DNA (Glover, White, Finnegan and Hogness, unpublished). Under conditions where the DNA on the filter is saturated with excess rRNA, we find that 2.3 and 4.2 kb of the Dml03 segment correspond to sequences in the 18 and 28S rRNAs respectively. Since these values are in good agreement with the lengths of these two rRNAs in *Drosophila* (2.2 and 4.2 kb; Loening, 1968), we conclude that Dml03 contains one copy of each of the corresponding sequences.

The repetition frequency of *c.* 400 copies per genome observed for Dml03 in our experiments with embryonic DNA is somewhat higher than the value of 250 rRNA genes per genome determined by Tartoff (1971) with DNA from adult flies. Although this difference is probably not significant (particularly since the data points for the *Drosophila* DNA driven reassociation of pDml03 are highly scattered), it may also result from a difference in the average gene frequencies for embryonic and adult cells (Spear and Gall, 1973).

The Dml03 segment is bounded by two EcoR1 termini and contains no internal EcoR1 cleavage sites. We suppose that this segment was cut out from the middle or from the end of a tandem array of repeating gene-spacer units like that found in *Xenopus* (Brown and Sugimoto 1974) and which apparently have been visualized as transcription matrices in *D. melanogaster* (Hankalo et al., 1974; Laird, personal communication) and *D. hydei* (Meyer and Hennig, 1974). If cut from the middle, we expect that the repeating unit should be equal to or greater than the length of Dml03 (17.1 kb), provided that sequence heterogeneity in the repeated unit does not involve the EcoR1 site(s). This would yield a spacer of $\geq c.$ 9 kb, given that the primary transcript has a length of about 8 kb in *Drosophila* (Greenberg, 1969; Meyer and Hennig, 1974). This is about twice the length of the spacer

in *Xenopus* (Brown and Sugimoto, 1974), and even a larger multiple of the spacer length estimated from the transcription matrices in *Drosophila* (however, these matrix estimates are quite inaccurate as they are highly dependent on spreading conditions; Meyer and Hennig, 1974; Laird, personal communication).

Were the Dml03 segment cut from the end of a tandem array, then the repeating unit, and hence the spacer, could be less than 17.1 and 9 kb, respectively. This possibility of an end cut is also of interest in regard to the *c.* 100-fold repetition in the Dml03 segment of small sequences of *c.* 30 base pairs. If the segment derives from a middle cut, these sequences would have to be assigned to the spacer region, whereas if it is derived from the end, they could represent sequences adjacent to the tandem array. This introduces the unanswered question of whether the 250 to 400 rRNA genes per nucleolus organizer (NO) in the X or Y heterochromatin (there is one NO in X and one in Y; Lindlsey and Grell, 1968) are in one long tandem array, or whether they are organized into many closely linked arrays of smaller size. Clearly if the latter topography prevails, the chance of isolating a segment from the end of an array is increased, and the repeated small sequences could be imagined to fill the gap between the linked arrays.

The Dml03 length could also be larger than the length of the average repeating unit if sequence heterogeneity in these units caused the EcoRI cleavage sites to be located at different loci in two adjacent units. A middle cut could then yield a segment in which part of the spacer region was duplicated. Such a small duplication would probably not be detected by the reassociation kinetics, given the presence of the more highly repeated component discussed above. Sequence heterogeneity of the spacers in rDNA in *Xenopus laevis* has been detected (Reeder, personal communication), and its presence in *D. melanogaster* could account for the scatter in the data points that we observed for the reassociation of pDml03 DNA in the presence of *D. melanogaster* DNA. This is because the ability of the S₁ endonuclease to hydrolyze DNA strands in regions containing mismatched base pairs is quite sensitive to slight changes in salt concentration and temperature (Shenk et al., this volume). Heterogeneity would produce such mismatches in the duplexes formed between the labelled Dml03 fragments and the unlabelled fragments from the *D. melanogaster* DNA, and these could cause the exceptional variation observed in the assay.

We conclude by calling attention to the selection that we accidentally employed to isolate the Dml03 segment, and the possible use of this selection for the isolation of other segments containing the rRNA gene. The combination of the sizing procedures for chromosomal segments and the use of the EcoRI-ligase joining method provided the selection. Fractionation for most other length classes of

chromosomal segments with EcoRI termini would have eliminated this segment. On the other hand, fractionation for 17 kb segments after *complete* EcoRI cleavage of the chromosomal DNA should enhance the selection for this segment; e.g., this procedure would eliminate all of the other Dm segments that we isolated by the EcoRI-ligase method. Indeed, the speculations given in the preceding paragraphs concerning the nature of the Dm103 segment can in large measure be tested by examining a set of Dm segments isolated by such an enhanced selection procedure.

ACKNOWLEDGMENT

This work was supported by research grants from the National Science Foundation and the National Institutes of Health. P.C.W. was a Dernham fellow and, more recently, a fellow of the American Cancer Society. D.M.G. is a Damon Runyon fellow, and R.L.W. and D.J.F. are Jane Coffin Child fellows. J.E.D. was a Helen Hay Whitney fellow.

REFERENCES

- Beermann, W. 1972. Chromomeres and genes. *In* Results and Problems in Cell Differentiation (W. Beermann, ed.) Vol. 4, p.1, Springer-Verlag, New York.
- Bishop, J.O. 1974. The gene numbers game. *Cell* 2: 81.
- Blumenthal, A.B., H.J. Kriegstein and D.S. Hogness. 1974. The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 205.
- Boyer, H.W. and D. Roulland-Dussoix. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41: 459.
- Britten, R.J. and E.H. Davidson. 1969. Gene regulation for higher cells: A theory. *Science* 165: 349.
- Brown, D.D. and K. Sugimoto. 1974. The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 501.
- Callan, H.G. 1967. The organization of genetic units in chromosomes. *J. Cell Sci.* 2: 1.
- Champoux, J.J. and D.S. Hogness. 1972. The topography of lambda DNA: Polyriboguanilylic acid binding sites and base composition. *J. Mol. Biol.* 71: 383.
- Clewell, D.D. and D.R. Helinski. 1969. Super coiled circular DNA-protein complex in *Escherichia coli*: Purification and induced conversion to an open circular DNA form. *Proc. Nat. Acad. Sci. USA* 62: 1159.
- Cohen, S.N., A.C.Y. Chang, H.W. Boyer and R.B. Helling. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Nat. Acad. Sci. USA* 70: 3240.

- Crick, F. 1971. General model for the chromosomes of higher organisms. *Nature* 234: 25.
- Davidson, E.H., B.R. Hough, C.S. Amenson and R.J. Britten. 1973. General interspersion of repetitive with nonrepetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* 77: 1.
- Davidson, N. and W. Szybalski. 1971. Physical and chemical characteristics of lambda DNA. In *The Bacteriophage Lambda* (A.D. Hershey, ed.) p. 45. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. In *Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 21D, p.413. Academic Press, New York.
- Georgiev, G.P. 1969. On the structural organization of operon and the regulation of RNA synthesis in animal cells. *J. Theor. Biol.* 25: 473.
- Graham, D.E., B.R. Neufeld, E.H. Davidson and R.J. Britten. 1974. Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. *Cell* 1: 127.
- Greenberg, J.R. 1969. Synthesis and properties of ribosomal RNA in *Drosophila*. *J. Mol. Biol.* 46: 85.
- Grigliatti, T.A., B.N. White, G.M. Tener, T.C. Kaufman, J.J. Holden and D.T. Suzuki. 1974. Studies on the transfer RNA genes in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 461.
- Hamkalo, B.A., O.L. Miller and A.H. Bakken. 1974. Ultrastructure of active eukaryotic genes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 915.
- Hedgpeth, J., H.M. Goodman and H.W. Boyer. 1972. DNA nucleotide sequence restricted by the R₁ endonuclease. *Proc. Nat. Acad. Sci. USA* 69: 3448.
- Hochmann, B. 1974. Analysis of a whole chromosome in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 581.
- Hogness, D.S. and J.R. Simmons. 1964. Breakage of λ dg DNA: Chemical and genetic characterization of each isolated half-molecule. *J. Mol. Biol.* 9: 411.
- Jackson, D.A., R.H. Symons and P. Berg. 1972. Biochemical method for inserting new genetic information into DNA of Simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* 69: 2904.
- Judd, B.H., M.W. Shen and T.C. Kaufman. 1972. The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics* 71: 139.

- Kato, K.-I., J.M. Goncalves, G.E. Houts and F.J. Bollum. 1967. Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase. *J. Biol. Chem.* 242: 2780.
- Kavenoff, R. and B.H. Zimm. 1973. Chromosome-sized DNA molecules from *Drosophila*. *Chromosoma* 41: 1.
- Kriegstein, H.J. and D.S. Hogness. 1974. Mechanism of DNA replication in *Drosophila* chromosomes: Structure of replication forks and evidence for bidirectionality. *Proc. Nat. Acad. Sci. USA* 71: 135.
- Laird, C.D. 1971. DNA of *Drosophila* chromosomes. *Ann. Rev. Genet.* 7: 177.
- Lee, C.S. and C.A. Thomas, Jr. 1973. Formation of rings from *Drosophila* DNA fragments. *J. Mol. Biol.* 77: 25.
- Lefevre, G. 1974. The one band-one gene hypothesis: Evidence from a cytogenetic analysis of mutant and nonmutant rearrangement breakpoints in *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 591.
- Lefevre, G. and M.M. Green. 1972. Genetic duplication in the white-split interval of the X chromosome in *Drosophila melanogaster*. *Chromosoma* 36: 391.
- Lindsley, D.L. and E.H. Grell. 1968. Genetic Variations of *Drosophila melanogaster*. Carnegie Institution, Washington, D.C.
- Little, J.W., I.R. Lehman and A.D. Kaiser. 1967. An exonuclease induced by bacteriophage λ : Preparation of the crystalline enzyme. *J. Biol. Chem.* 242: 672.
- Lobban, P.E. and A.D. Kaiser. 1973. Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* 78: 453.
- Loening, V.E. 1968. Molecular weights of ribosomal RNA in relation to evolution. *J. Mol. Biol.* 38: 355.
- Mandel, J. and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53: 159.
- Marvin, D.A. and B. Hohn. 1969. Filamentous bacterial viruses. *Bact. Rev.* 33: 172.
- Mertz, J.E. and R.W. Davis. 1972. Cleavage of DNA by R₁ restriction endonuclease generates cohesive ends. *Proc. Nat. Acad. Sci. USA* 69: 3370.
- Meyer, G.F. and W. Hennig. 1974. The nucleolus in primary spermatocytes of *Drosophila hydei*. *Chromosoma* 46: 121.
- Morrow, J.F. 1974. Mapping the SV40 chromosome by use of restriction endonucleases. Ph.D. thesis, Stanford University.
- Morrow, J.F., S.N. Cohen, A.C.Y. Chang, H.W. Boyer, H.M. Goodman and R.B. Helling. 1974. Replication and transcription of eukaryotic DNA in *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* 71: 1743.
- O'Brien, S.J. 1973. On estimating functional gene number in eukaryotics. *Nature New Biology* 242: 52.

- Paul, J., R.S. Gilmour, N. Affara, G. Birnie, P. Harrison, A. Hell, S. Humphries, J. Windass and B. Young. 1974. The globin gene: Structure and expression. *Cold Spring Harbor Symp. Quant. Biol.* 38: 885.
- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 405.
- Prescott, D.M. and K.G. Murti. 1974. Chromosome structure in ciliated protozoans. *Cold Spring Harbor Symp. Quant. Biol.* 38: 609.
- Radloff, R., W. Bauer and J. Vinograd. 1968. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. *Proc. Nat. Acad. Sci. USA* 57: 1514.
- Rasch, E.M., H.J. Barr and R.W. Rasch. 1971. The DNA content of sperm of *Drosophila melanogaster*. *Chromosoma* 33: 1.
- Rhoades, M., L.A. MacHattie and C.A. Thomas, Jr. 1968. The P22 bacteriophage DNA molecule. II. Circular intracellular forms. *J. Mol. Biol.* 37: 21.
- Rudkin, G.T. 1961. Cytochemistry in the ultraviolet. *Microchem. J. Symp. Ser.* 1: 261.
- _____. 1972. Replication in polytene chromosomes. In Results and Problems in Cell Differentiation (W. Beermann, ed.) Vol. 4 p.59. Springer-Verlag, New York.
- Schachat, F.H. and D.S. Hogness. 1974. Repetitive sequences in isolated Thomas circles from *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 371.
- Sharp, P.A., B. Sugden and J. Sambrook. 1973. Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose. *Biochemistry* 12: 3055.
- Spear, B. and J.G. Gall. 1973. Independent control of ribosomal gene replication in polytene chromosomes of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 70: 1359.
- Tartoff, K.D. 1971. Increasing the multiplicity of ribosomal RNA genes in *Drosophila melanogaster*. *Science* 171: 294.
- Thomas, C.A. Jr. 1970. The theory of the master gene. In The Neurosciences: Second study program (F.O. Schmitt, ed.) p.973. The Rockefeller University Press, New York.
- Thomas, M. and R.W. Davis. 1975. Studies on the cleavage of bacteriophage lambda DNA with EcoRI restriction endonuclease. *J. Mol. Biol.* In press.
- Wensink, P.C., D.J. Finnegan, J.E. Donelson and D.S. Hogness. 1974. A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. *Cell*. In press.
- Wimber, D.E. and D.M. Steffensen. 1970. Localization of 5S RNA genes on *Drosophila* chromosomes by RNA-DNA hybridization. *Science* 170: 639.

Sequences of highly repeated DNA in *Drosophila melanogaster*

D. L. Brutlag and W. J. Peacock

Most of the highly repeated DNA of *Drosophila melanogaster* is composed of four species with buoyant densities 1.672, 1.686, 1.688 and 1.705 g/cc in CsCl. The 1.672 satellite contains pentameric (ATAAT) and heptameric (ATATAAT) repeats. Several lines of data indicate the pentamers and heptamers are on different homogeneous molecules. Related sequences (isomers) are also found in the 1.686 satellite, the major repeat unit being AATAACATAG. A similar sequence (AGAGAAGAAG) is predominant in the 1.705 species. We have no evidence for sequence heterogeneity caused by random mutation, and we suggest that other cases of apparent sequence heterogeneity may resolve to be cases of sequence isomers. We propose that the highly repeated DNAs of any one species will show marked relatedness and that their arrangement may be important in homologous chromosome recognition.

The highly repeated DNA fractions of many higher organisms often resolve as satellite peaks in equilibrium density gradients. The variable amounts of satellite DNA and their diverse buoyant densities between closely related species have led to the suggestion that they are not conserved over even short periods of evolutionary history (Hennig et al., 1969; Hennig and Walker, 1970). The location of many satellites within the centromeric constitutive heterochromatin indicates that these sequences are inactive (not transcribed) and is consistent with them playing a structural and not an informational role in the genome.

Sequence analysis of a number of satellite DNAs (crab, Swartz et al., 1962; guinea pig, Southern, 1970; *Drosophila virilis*, Gall and Atherton, 1974; kangaroo rat, Fry et al., 1973) has revealed very short repeating units (2 to 10 nucleotides long), further reducing the possibility of a coding function. These analyses have also shown that the evolutionary stability of satellite DNA may be greater than has been supposed. In a short nucleotide sequence, a single base change can result in a large change in buoyant density. For example, the three satellites of *D. virilis* have different densities but each contains a homogeneous repeating heptanucleotide which differs from the other two

satellites by only one or two base substitutions (Gall and Atherton, 1974). In addition to the homogeneity and similarity of satellite sequences within the species, Gall and Atherton found that a satellite common to two sibling species (*D. virilis* and *D. americana*) had an identical seven nucleotide repeating unit in each species. Hybridization experiments, although not as sensitive as direct sequence analysis, have also detected homologies between satellites of closely related species (*D. virilis*, Blumenfeld et al., 1973; mouse, Sutton and McCallum, 1972; Rice and Strauss, 1973; primates, Prosser et al., 1973). These data argue that satellite DNA sequences are conserved and that they are subject to selection.

In contrast to the evidence for homogeneous repeating units in the *D. virilis* satellites, several experiments have suggested sequence heterogeneity. Renatured satellite DNA often shows a large number of mismatched nucleotides as inferred both from altered buoyant density and reduced thermal stability (Sutton and McCallum, 1971). Sequence analyses have also suggested the presence of nucleotide alterations. Most of the minor pyrimidine tracks present in crab and guinea pig satellite DNAs are consistent with single base substitutions or insertions in the major repeating unit (Ehrlich et al., 1973; Southern, 1970). The heterogeneity present in the kangaroo rat satellite consists of a number of sequences related by single base insertions (Fry et al., 1973). Sequence heterogeneity has been interpreted to be the result of random mutation, implying that exact sequence in these DNAs may not be of critical importance for their function.

We have purified the highly repeated satellite DNA fractions from *Drosophila melanogaster* and have begun an analysis of their nucleotide sequence. The sequence heterogeneity present in at least two of the *D. melanogaster* satellites is clearly not due to random mutation. We have found two or more closely related sequences within each satellite species, and in at least one of the satellites the related sequences are on different DNA molecules and perhaps on different chromosomes. These data suggest that highly repeated DNA may be involved in cellular processes requiring homologous chromosome recognition.

RESULTS

D. melanogaster contains four discrete satellite fractions The highly repeated DNA of *D. melanogaster* can be isolated as satellite peaks in CsCl gradients containing actinomycin-D or netropsin sulphate (Peacock et al., 1974). Four distinct density species (1.672, 1.686, 1.688 and 1.705 g/cc) can be isolated, each comprising 3-5% of the DNA. These four satellites account for 85% of

the highly repeated DNA of the genome. Each satellite appears homogeneous in neutral CsCl and each separates into two complementary strands in alkaline CsCl. While each appears to be a single buoyant density species, thermal denaturation profiles and hybridization experiments indicate small amounts of cross contamination (up to 5%) between the various satellites (Goldring et al., 1975). Although these contaminants have been shown to be due to covalent attachment of one repeated sequence to another, they can be reduced in amount by repurifying each satellite in a second gradient, or by isolating the complementary single strands in alkaline gradients.

Sequence analysis of the 1.672 satellite indicates that it consists of two different DNA species To determine the basic repeating sequence of the satellites we have used either repurified native DNA or isolated complementary strands as templates for *Escherichia coli* RNA polymerase or *E. coli* DNA polymerase I. Radioactive DNA or RNA was synthesized (c-DNA or c-RNA) and then sequenced by standard procedures.

When the labeled 1.672 DNA was subjected to pyrimidine tract analysis (to determine runs of pyrimidine bases) only two major pyrimidine oligonucleotides were found, T₂ and T₁ (molar ratio, 1:3.24). This satellite therefore consists predominantly (> 96%) of the subunits $\begin{matrix} \text{TTA} \\ \text{AA} \end{matrix}$ and $\begin{matrix} \text{TA} \\ \text{AT} \end{matrix}$ (Peacock et al., 1974).

Sequence analysis of RNA complementary to the heavy strand of 1.672 DNA agrees with this formulation and allows an ordering of the di- and tri-nucleotides. Pancreatic ribonuclease digests give only two major products, AAU and AU, in the molar ratio 1:1.43. To order these short oligonucleotides, we analysed partial pancreatic ribonuclease digests where longer products are still present (Table 1). In addition to AAU and AU, most of the radioactivity was found in two major partial products - a pentanucleotide composed of one AAU and one AU, and a heptanucleotide containing one AAU and two AUs. Analysis of these two oligonucleotides by stepwise degradation with spleen phosphodiesterase demonstrated that all possible isomeric forms with the given composition were present in the digests (i.e. AUAAU and AAUAU for the pentameric sequence, and AUAUAAU, AUAAUAU, and AAUAUAU for the heptamer). Since the penta- and hepta-nucleotides are the majority of partial products they may represent basic repeating units in the 1.672 DNA.

The presence or absence of certain of the oligonucleotides in the series adds further support to the evidence for penta- and heptanucleotide repeats, and yields information as to their arrangement. Hexanucleotides were absent from the series of oligonucleotides of

TABLE 1

Oligonucleotides present in partial pancreatic ribonuclease digests of RNA complementary to 1.672 heavy strand

Length (in nucleotides)	Oligonucleotide composition	
	Inferred	Determined (AU:AAU)
2	(AU) ₁	AU (> 95%)
3	(AAU) ₁	AAU (> 95%)
4	(AU) ₂	AU (> 95%)
5	(AU) ₁ (AAU) ₁	1.07
6	-	-
7	(AU) ₂ (AAU) ₁	2.14
8	(AU) ₁ (AAU) ₂	0.7
9	(AU) ₃ (AAU) ₁	2.72
10	(AU) ₂ (AAU) ₂	1.1
11	(AU) ₄ (AAU) ₁	3.1
12	(AU) ₃ (AAU) ₂	1.65
13	(AU) ₂ (AAU) ₃	0.96
14	(AU) ₄ (AAU) ₂	2.34
15	(AU) ₃ (AAU) ₂	1.24
16	(AU) ₅ (AAU) ₂	N.A.
17	(AU) ₄ (AAU) ₃	N.A.
18	(AU) ₃ (AAU) ₄ or (AU) ₆ (AAU) ₂	N.A.
19	(AU) ₅ (AAU) ₃	N.A.
20	(AU) ₄ (AAU) ₄	N.A.

³²P-RNA annealed to its DNA template (1.672 heavy strand) was partially digested with pancreatic ribonuclease and the products separated by electrophoresis on cellulose acetate, pH 3.5, followed by homochromatography (Brownlee, 1972). Both the length and base composition are determined by the position of the oligonucleotide on the chromatogram. The oligonucleotide composition could be inferred from the base composition and the length, and from the fact that this RNA contains only two major pancreatic products, AU and AAU. These assignments have been verified by elution of oligonucleotides and complete pancreatic digestion. Partial products longer than 15 residues were not analyzed (N.A.).

the partial digest. Since neither $(AAU)_2$ nor $(AU)_3$ was generated, it follows: (i) two AAUs are never adjacent to one another and must be separated by at least one AU dinucleotide, and (ii) that there may be only one or two but not three AUs between successive AAUs. These rules define pentameric or heptameric sequences.

Longer oligonucleotides indicate the arrangement of these two units relative to each other. The oligonucleotide $(AU)_4(AAU)_1$ could only arise if two of the AU dinucleotides were adjacent. Since two AAUs are never next to each other, the presence of $(AU)_2(AAU)_3$ indicates that two pentanucleotide repeats must also be adjacent. Similarly the presence of the two 18-mer oligonucleotides $(AU)_3(AAU)_4$ and $(AU)_6(AAU)_2$ implies that three of the pentanucleotides are adjacent and that three of the heptanucleotides are adjacent. This pattern of partial products suggests that the 1.672 satellite might contain two different DNA species, one with a five nucleotide repeated sequence and the other with a seven nucleotide repeat.

Thermal denaturation indicates two DNA species in the 1.672 satellite This sequence information predicts that isolated complementary strands from either a pentamer or heptamer repeat could self-anneal in a hairpin type DNA molecule. If an AATATAT repeating sequence were to anneal with itself there would be one A-A mismatch every seven nucleotides. Similarly, the pentanucleotide repeat could self-anneal with one mismatch in every fifth base pair. Examination of isolated heavy strand DNA by electron microscopy shows that most is in the form of branched double stranded molecules (Figure 1). Thermal denaturation of the purified complementary strands shows two transitions at temperatures consistent with 20% and 14% mismatched nucleotides (Figure 2). A third transition at the T_m of native 1.672 DNA indicates minor contamination of each strand by its complement. These data confirm that the predicted self-annealing does occur and indicate that there are long regions of pentameric repeats and long regions of heptameric repeats.

DNA sequence analysis indicates that these two different repeating units occur in different DNA molecules. Either complementary strand of 1.672 serves as an excellent template for *E. coli* DNA polymerase in the absence of any added primer DNA. Since DNA polymerase requires a primer annealed to a template, this is further evidence for the double stranded nature of these molecules. Analysis of pyrimidine tracts present in DNA complementary to the light strand shows predominantly T_2 and T_1 in the ratio 1:1.42, in good agreement with the AAU:AU ratio determined in the c-RNA analysis described in the previous section. If the light strand is

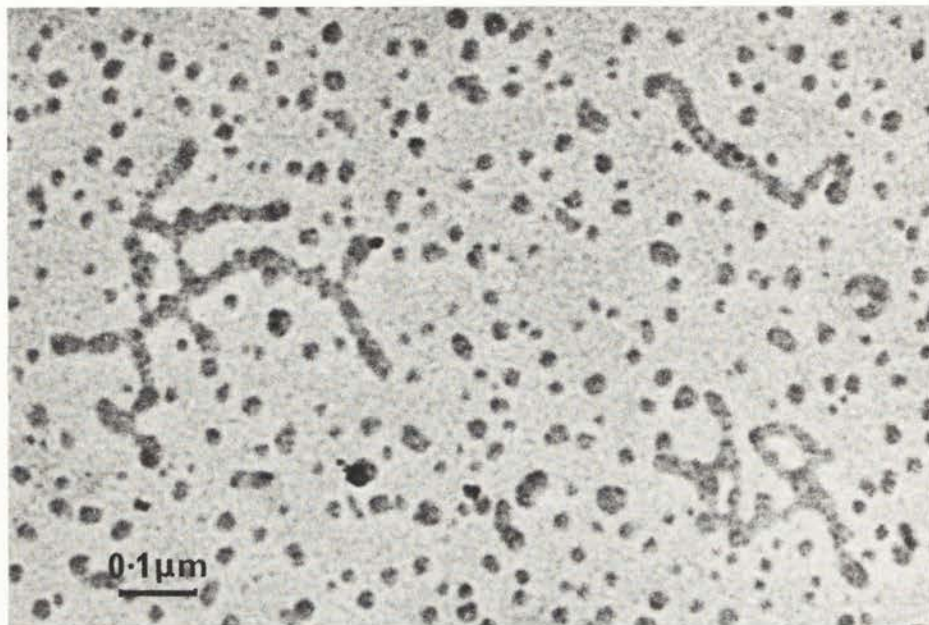


Figure 1. Electron micrograph of molecules of the heavy strand of the 1.672 satellite. The molecules were isolated in preparative alkaline CsCl gradients (Peacock et al., 1974) and prepared for electron microscopy by the method of Davis et al., 1971.

used as template for DNA polymerase in the absence of dGTP and dCTP, DNA is still synthesised (45% of the amount synthesised with all four triphosphates present). The product, however, contains $T_2:T_1$ in the ratio 1:1.03, indicating that only the pentanucleotide repeat is being replicated. The heptanucleotide sequence would give T_2 and T_1 in a 1:2 ratio. This experiment shows that the few G and C residues in the 1.672 satellite must be located predominantly in the heptanucleotide peak fraction since the absence of these nucleotides prevents DNA polymerase from replicating the heptamer. The amount of pentanucleotide repeat synthesised seems unaffected by the absence of dGTP and dCTP. The length of the template (3 kb), and the amount of DNA synthesised (6.1% of template), indicates that each DNA polymerase molecule synthesises 180 nucleotides on the average. Therefore the pentanucleotide repeat must contain less than one G or C residue per 180 nucleotides. An upper estimate for the frequency of G residues of one per 400 nucleotides was obtained by T_1 -ribonuclease digestion of complementary RNA.

The pentanucleotide repeat is therefore very homogeneous and we are able to exclude the possibility of it containing any random mutations over stretches of many hundreds of nucleotides.

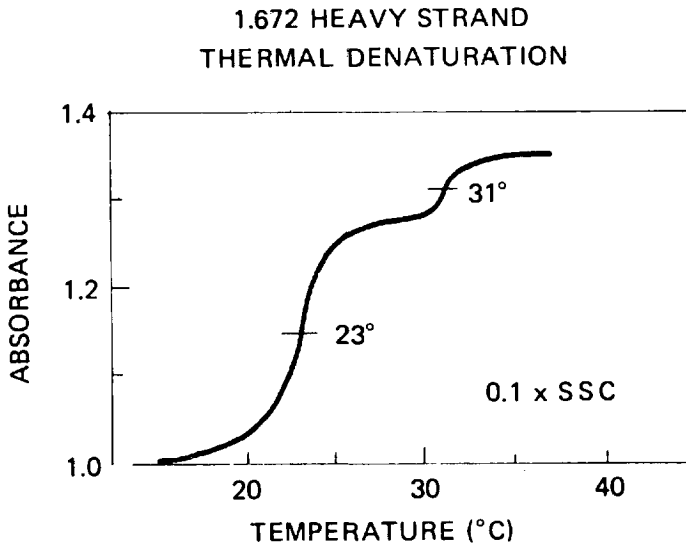


Figure 2. Melting profile of the heavy strand of the 1.672 satellite. The heavy strand of the 1.672 satellite was isolated in preparative alkaline CsCl gradients (Peacock et al., 1974) and dialyzed extensively against 0.1 x SSC (pH 7.3). The DNA was then heated (0.5%/min) and the absorbance recorded automatically.

Moreover, those molecules which do serve as templates in the absence of G and C are composed entirely of the pentanucleotide repeat, implying that the two different repeats are on different molecules. We have no information on the distribution of G and C bases in the heptanucleotide repeat, but if they were regularly dispersed, a higher order regularity within this DNA would be indicated. A long order organization has recently been demonstrated in a highly repeated DNA. Botchan (1974) found a 1400-nucleotide repeat in bovine satellite I, which was itself internally repetitive, and Thomas et al. (1975) have evidence of a supra-organization of repeats in the 1.688 satellite of *D. melanogaster*.

The 1.686 satellite also contains more than one repeat sequence Unlike the AT-rich 1.672 satellite, the 1.686 DNA species contains about 10% G and 10% C residues. T₁-ribonuclease digestion of RNA complementary to the heavy strand of this DNA yields predominantly oligonucleotides 10 residues long. The G bases are therefore evenly spaced every 10 nucleotides in this c-RNA. One major oligonucleotide has the composition A₆U₂C₁G₁, and the two minor components have the

compositions $A_7U_1C_1G_1$ and $A_5U_3C_1G_1$. Pancreatic ribonuclease cleavage of the major oligonucleotide gives the subunits AU, AAU, AAC and AG, in equimolar amounts. Stepwise degradation with spleen phosphodiesterase indicates the sequence of this oligonucleotide to be 5' AAUAACAUAG 3'. We have not analysed the sequence of the two minor oligonucleotides; however, pancreatic digestion indicates that their sequences are similar to the major one. Thus, like the 1.672 species, this satellite contains more than one closely related repeated sequence. The different sequences found in the 1.686 DNA are probably not each on a different DNA molecule because the sequences have different enough base compositions to resolve as separate peaks in alkaline CsCl gradients. However, the possibility that the major repeating unit is present on one molecule of DNA and the two minor ones on another molecule, cannot be excluded. Two such DNAs would be sequence isomers and should have identical properties in both neutral and alkaline gradients. The 1.686 sequences differ from the repeats in the 1.672 satellite in a number of respects: the length of the repeat unit, the presence of G and C residues, and the adjacent location of two A-A-pyrimidine trinucleotides. The two satellites are similar in that each possesses an A-rich and a T-rich strand, and in the A-rich strand the T, G and C bases are separated from each other by either one or two A residues.

The 1.705 satellite may also have a simple, related sequence

This satellite contains a purine-rich strand and a pyrimidine-rich complementary strand. Because RNA polymerase must initiate with ATP or GTP, it preferentially synthesizes RNA complementary to only the pyrimidine-rich strand. The enzyme can in fact synthesize c-RNA, with native 1.705 DNA as template, in the absence of both UTP and CTP. T_1 -ribonuclease digestion of the RNA product from a synthesis containing only ATP and GTP, gives AG and AAG oligonucleotides in equimolar amounts. The same ratio of these two oligonucleotides was produced by T_1 -ribonuclease digestion of c-RNA made in the presence of all four triphosphates, but in addition, a few minor oligonucleotides containing U and C were found. While a complete sequence analysis has not been performed, one major partial digestion product is an oligonucleotide 10 residues long. The sequence of this oligonucleotide, 5' AGAGAAGAAG 3', is closely related to the major 1.686 repeat (5' AUAGAAUAC 3'), differing by three base substitutions.

Recently, Birnboim, Straus and Sederoff (personal communication) have found very long runs of pyrimidines in *D. melanogaster* DNA (350 nucleotides), constituting 4.2% of the genome. They suggested that the tracts could represent the pyrimidine-rich strand of the

1.705 satellite. To test this hypothesis we isolated the pyrimidine tracts from purified 1.705 DNA and found them to be all eluted in the excluded volume of a G75 sephadex column. Moreover the sequence of RNA complementary to the long pyrimidine tracts of total nuclear DNA is similar to that of 1.705 c-RNA; the T₁-ribonuclease digests of both are almost identical, with the major oligonucleotides being AG and AAG.

CONCLUSIONS

Relatedness of satellite DNAs The three satellite DNAs of *Drosophila melanogaster* examined here (1.672, 1.686, 1.705 g/cc) are similar to each other in containing short repeat units (5, 7 and 10 nucleotides long). These repeat sequences are related to each other, differing by single base substitution in, and rearrangement of, di-(AT) and tri-nucleotide (AAT) units. The relationships between the satellites of *D. melanogaster* are more complex than those between the satellites of *D. virilis* where the only differences are base substitutions within a repeating heptanucleotide (Gall and Atherton, 1974). The *D. melanogaster* sequences appear to be quite unrelated to the satellite sequences of *D. virilis*. However, the relatedness between the various satellites within each of the species may prove to be a general case for any species with more than one satellite. Intraspecific satellite sequence similarity may be a requirement of the function of highly repeated DNA.

Sources of sequence heterogeneity Sequence analyses of the *D. melanogaster* satellites have disclosed some features of highly repeated DNA which are pertinent to apparent sequence heterogeneity. Initial studies of the sequence of the 1.672 satellite showed a high amount of G and C nucleotides (7-10%) in native DNA, and a nearest neighbour analysis indicated that these residues occurred in a regular pattern within the 1.672 repeating unit. However, purification of the complementary strands resulted in the loss of most of these residues (to less than 1% G or C). We therefore concluded that most of these nucleotides were not distributed throughout the 1.672 DNA, but were contained in a contaminating repeated DNA. Since the different satellite sequences appear to be covalently joined to one another as a result of their adjacent location within chromosomal DNA (Goldring et al., 1975), the presence of contaminants consisting of related repeated sequences may be unavoidable, and may often give the appearance of single base alterations within a satellite DNA.

A second possible cause of apparent sequence heterogeneity, usually indicated by lowered melting temperatures or increased buoyant densities of renatured DNA relative to native DNA, is the

existence of more than one repeated sequence within an apparently homogeneous satellite. The related sequences present in the 1.672 and 1.686 satellites result in the production of mismatched nucleotides during renaturation. If isomeric or other closely related sequences were present on different DNA molecules, or on the same DNA molecule in a regular arrangement, renaturation involving interaction of the different basic repeats would result in regular nucleotide mismatches.

Thus even if it is known that mismatched nucleotides occur in a renatured satellite it cannot be assumed that they are due to random mutations. The distribution of the mismatches should be examined to determine the nature of the sequence heterogeneity.

Homogeneity and a possible function of satellite sequences At least two of the *D. melanogaster* satellites appear to be homogeneous by sequence analysis. The absence of G and C residues in the pentanucleotide fraction of the 1.672 species suggests the absence of random mutations over long regions. The recovery of long pyrimidine tracts from 1.705 DNA and the simplicity of the T₁-ribonuclease digests indicate that a short repeat is conserved over long regions. Because of the presence of all four nucleotides in the basic repeat of the 1.686 satellite, a measure of its homogeneity will not be as straightforward.

The homogeneity of repeats argues that both the nucleotide sequence and the arrangement of the repeating units in satellite DNA carry information essential to its function.

Evidence that the 1.672 satellite is located in only two chromosomes (Goldring et al., 1975), and the fact that this satellite contains only two related sequences, suggest that highly repeated DNAs are distributed in a chromosome specific manner. Satellite DNAs may be important in cellular processes requiring homologous chromosome recognition. This hypothesis suggests that a single satellite which is found in many chromosomes (such as mouse satellite) may derive its sequence heterogeneity from the presence of a family of closely related sequences (each homogeneous) arranged to give chromosomal specificity.

REFERENCES

- Blumenfeld, M., A.S. Fox and H.S. Forrest. 1973. A family of three related satellite DNAs in *Drosophila virilis*. *Proc. Nat. Acad. Sci. USA* 70: 2772.
- Brownlee, G.G. 1972. Determination of sequences in RNA. *In Laboratory Techniques in Biochemistry and Molecular Biology* (T.S. Work and E. Work, ed.) North-Holland Publishing Co., Amsterdam.

- Botchan, M.R. 1974. Bovine satellite I DNA consists of repetitive units 1400 base pairs in length. *Nature* 251: 288.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *In Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 21D, p413.
- Ehrlich, S.D., J.P. Thiery and G. Bernardi. 1973. Pyrimidine tracts of the (A-T) rich satellite DNA from *Cancer paguru*. *Biochim. Biophys. Acta* 312: 633.
- Fry, K., R. Poon, P. Whitcome, J. Idriss, W. Salser, J. Mazrimas and F. Hatch. 1973. Nucleotide sequence of MS- β satellite DNA from kangaroo rat *Dipodomys ardi*. *Proc. Nat. Acad. Sci. USA* 70: 2642.
- Gall, J.G. and D.D. Atherton. 1974. Satellite DNA sequences in *Drosophila virilis*. *J. Mol. Biol.* In press.
- Goldring, E.S., D.L. Brutlag and W.J. Peacock. 1975. Arrangement of the highly repeated DNA of *Drosophila melanogaster*. This volume.
- Hennig, W., I. Hennig and H. Stein. 1970. Repeated sequences in the DNA of *Drosophila* and their localization in giant chromosomes. *Chromosoma* 32: 31.
- Hennig, W. and P.M.B. Walker. 1970. Variations in the DNA from two rodent families (Cricetidae and Muridae). *Nature* 225: 915.
- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 405.
- Prosser, J., M. Moar, M. Bobrow and K.W. Jones. 1973. Satellite sequences in chimpanzee (*Pan troglodytes*). *Biochim. Biophys. Acta* 319: 122.
- Rice, N.R. and N.A. Straus. 1973. Relatedness of mouse satellite DNA to DNA of various *Mus* species. *Proc. Nat. Acad. Sci. USA*. 70: 3546.
- Southern, E.M. 1970. The sequence and evolution of guinea pig α -satellite DNA. *Nature* 227: 794.
- Sutton, W.D. and M. McCallum 1971. Mismatching and the reassociation rate of mouse satellite DNA. *Nature New Biol.* 232: 83.
- _____, _____. 1972. Related satellite DNA's in the genus *Mus*. *J. Mol. Biol.* 71: 633.
- Swartz, M.N., T.A. Trautner and A. Kornberg. 1962. Enzymatic synthesis of DNA. XI. Further studies on nearest neighbor base sequences in DNA. *J. Biol. Chem.* 237: 1961.
- Thomas, C.A. Jr, D.H. Hamer, R.E. Pyeritz, G.P. Holmquist, K. Shastri and S. Manteiul. 1975. An analysis of *Drosophila* DNA by restriction endonucleases. This volume.

Arrangement of the highly repeated DNA of *Drosophila melanogaster*

E. S. Goldring, D. L. Brutlag and W. J. Peacock

The highly repeated DNA of *Drosophila melanogaster* occurs in a few long stretches (600 kb) in the chromosome complement. Cytological hybridization to polytene chromosomes shows that each of the highly repeated species is localized in the chromocentre and also at a site on the left arm of chromosome 2 (region 21C-D). The 1.686 and 1.705 density species hybridize to all chromosomes, but the 1.672 species is restricted to chromosome 2 and possibly also to the Y chromosome. Hybridization to mitotic chromosomes from tissue culture cells confirms the centromeric location of these highly repeated DNA species. Analysis of molecules containing the 'ends' of 1.705 sequences indicates that these segments are, in some chromosomes, adjacent to 1.686 segments, and in fewer chromosomes to 1.672 segments. We suggest that each chromosome may have a specific arrangement of highly repeated DNA sequences.

The highly repeated DNA of the *Drosophila melanogaster* genome can be isolated by buoyant density centrifugation. Four density species, accounting for about 85% of the highly repeated sequences, can be separated as satellite bands distinct from the bulk of nuclear DNA by the successive use of actinomycin-CsCl and netropsin-CsCl gradients. After removal of the antibiotics these satellites have CsCl buoyant densities of 1.672, 1.686, 1.688 and 1.705 g/cc (Peacock et al., 1974). As well as determining the base sequence of the repeating units of the satellite (Brutlag and Peacock, 1975), we have investigated their organization in the genome. In particular we have examined:

- (i) the length and number of segments in which each satellite occurs as a 'pure' species;
- (ii) the location of each satellite in the chromosome complement; and
- (iii) the arrangement of the various satellites relative to each other.

LENGTH OF THE SATELLITE SEGMENTS

Each satellite comprises about 4% of the genome, or approximately 6,000 kilobases (kb); this DNA could be in one long segment

or there could be segments at more than one locus in the chromosomal complement. If the mean chromosomal segment length of a particular satellite were smaller than the mean size of the isolated DNA molecules, then many of the satellite sequences would not band at the satellite density in a gradient. Yields of the 1.705 satellite were equivalent from populations of molecules of 2.4 and 30.0 kb mean size, implying a mean chromosomal segment length greater than 30 kb (Peacock et al., 1974). In both gradients only 5% of the 1.705 DNA remained in the main band region. This suggested that the chromosomal segment length of the 1.705 species was of the order of 20 x 30 kb.

We have extended this analysis using tissue culture cells from which long DNA molecules are readily obtained. An analysis of the relationship between yield and molecular length was possible for the 1.705 and 1.686 satellites because of their positions in actinomycin-CsCl and netropsin-CsCl gradients. In actinomycin-CsCl, the 1.705 satellite is the most dense band and is clearly separated from the next band (1.686 + 1.672) and from the lighter main band region. In netropsin-CsCl the 1.686 DNA is the least dense band. As the length of the molecules increases we can expect that there will be an increase in the proportion of molecules containing more than one sequence type, i.e. different sequences are more likely to be included in any one molecule. Such mixed-sequence molecules will take a position in the gradient dependent upon the relative amounts of each sequence. A density peak, which with short lengths of DNA is pure satellite, could, with long DNA, be composed of different molecule types. In choosing the 1.705 and 1.686 satellites we avoided this form of sequence contamination. When molecule length approaches the segment length of these two satellites, the amounts banding at their characteristic densities will be reduced. We have compared the yields of these satellites with DNA sheared to short lengths (4.5 kb) or isolated as very long molecules (225 kb). The use of two different isotopes enabled this comparison to be made in a single gradient; the details of methods and results are given in Table 1. The experiment showed that when the mean molecule length was 225 kb the 1.705 peak was reduced only 17% relative to the sheared control. This implies that the 1.705 satellite occurs in segments that are in the range of 225 kb to 6 x 225 kb in length. It is difficult to be more precise than this until the sequences neighbouring the 1.705 segments have been defined - the buoyant density shifts of mixed molecules obviously depend upon sequence composition. In another experiment where the molecules were greater than 270 kb, the 1.705 peak was reduced by about 20%.

Both of these experiments suggest that the earlier estimate (Peacock et al., 1974) of segment length of 600 kb is of the right

TABLE 1

Satellite recoveries from sheared and unsheared DNA

(a) Actinomycin D-CsCl gradient

³H labelled DNA - average length 225 kb

¹⁴C labelled DNA - average length 4.5 kb

	Counts in 1.705 peak	Counts in 1.686 and 1.672 peak	Total counts in gradient
³ H	87,671 (3.8%)	123,204 (5.4%)	2.29 x 10 ⁶
¹⁴ C	4,291 (4.6%)	5,171 (5.5%)	93,277

(b) Netropsin-CsCl gradient

³H labelled DNA - average length 180 kb

¹⁴C labelled DNA - average length 4.5 kb

	Counts in 1.686 peak	Total counts in gradient
³ H	34,407 (2.1%)	1.67 x 10 ⁶
¹⁴ C	2,532 (2.0%)	1.28 x 10 ⁵

Tissue culture cells (line 2, Schneider, 1972) in log phase growth were labelled by the addition of 100 μ Ci of ³H-thymidine (5-Me T, 5Ci/mM), or by the addition of 5 μ Ci of ¹⁴C-thymidine (62 mCi/mM) for 4 hr. The cells were harvested by centrifugation (2,000 rpm, 10 min), washed in 0.79% NaCl and resuspended in buffer (0.25 M sucrose, 0.03 M Tris, 1 mM EDTA, 2.5 mM CaCl₂ pH 7.5). Sarkosyl was added to 0.5%, and saturated CsCl layered underneath the sample. Actinomycin or netropsin was added to a final concentration of 40 μ g/ml, and the density adjusted to 1.66 g/cc for actinomycin-D gradients or 1.62 g/cc for netropsin gradients. The tubes were centrifuged for 60 hr at 39,000 rpm in a #40 rotor (Spinco). Fractions were collected and 50 λ aliquots spotted onto glass fibre discs, dried and counted in a Packard liquid scintillation counter. The two channels were set so that the spillover from the ³H channel into the ¹⁴C channel was negligible, and from ¹⁴C into ³H, 10%. The ratio of ³H to ¹⁴C counts was 10:1. The ¹⁴C-DNA was sheared, prior to CsCl addition, through a 26 gauge needle. Samples were withdrawn and spread immediately for electron microscopy by the method of Davis et al. (1971). A mass average molecular weight was calculated. If ¹⁴C and ³H labelled DNAs were sheared together, the proportion of DNA in the 1.705 band was found to be the same for both samples.

order and that there must only be relatively few chromosomal locations of the 1.705 sequences since the total length of 1.705 DNA in the genome is only 6,000 kb. The 1.686 satellite must also occur in very few segments in the genome since 4.5 kb and 180 kb DNA gave the same yields of this species. We are unable to study either the 1.672 or 1.688 satellites by this method; 1.672 is virtually absent from these tissue culture cells and the 1.688 peak, in both

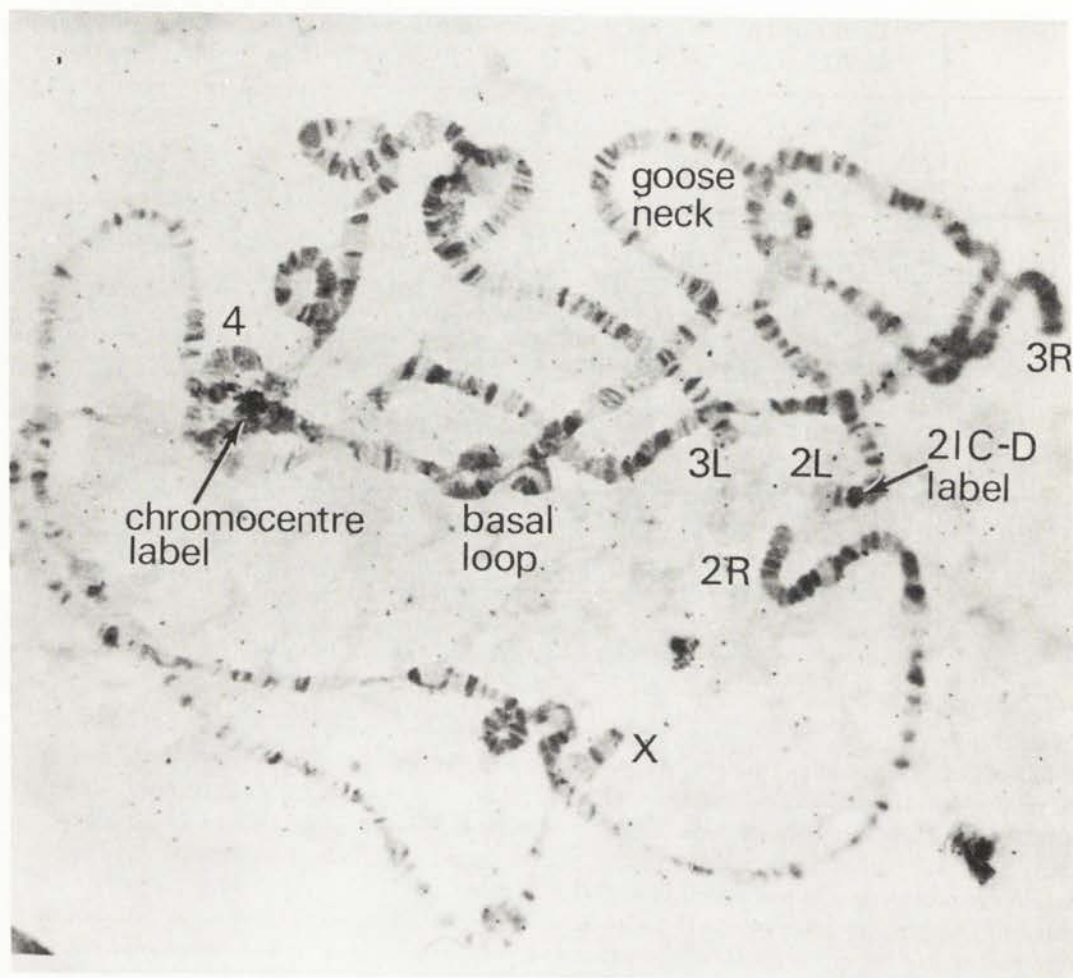


Figure 1. Cytological hybridization to salivary gland chromosomes. ^3H -RNA was synthesized using the heavy strand of the 1.686 satellite as template (after Pardue and Gall, 1970). Chromosomal DNA in salivary gland squashes was denatured in 0.2 N HCl, 30 min at 37°C. ^3H -RNA (3×10^5 cpm in 3 λ) was hybridized by heating to 90°C for 15 min and then at 66°C for 16 hr. After RNase digestion, slides were dipped in NTB-2 (Kodak) or L-4 (Ilford) emulsion. Exposure was usually 4-7 days and slides were then stained in Giemsa.

actinomycin-CsCl and netropsin-CsCl gradients, is subject to contamination by main band and mixed sequence molecules. Preparative gradients have provided some information about these two satellites; the yields of both were unchanged in gradients of 4.5 kb or 30 kb DNA. Therefore it is probable that all four highly repeated DNA species occur in a limited number of segments in the genome.

LOCATION OF THE SATELLITE SEGMENTS

Cytological hybridization (Pardue and Gall, 1970) confirms that the satellites are present in only a few sites in the chromosome complement. Tritiated RNAs synthesized using the satellite DNAs as templates, when hybridized to the polytene chromosomes in salivary gland cells, localize to the chromocentre (aggregated centromeres) and to region 21C-D on the left arm of chromosome 2 (Figure 1). This could mean that each satellite occurs in the centromeric region of each chromosome. Occasionally, individual chromosomes separate from the chromocentre during the squash procedure, and these cells yield more specific information about the chromosomal location of the satellites. Data from such cells (Figure 2) indicate that the 1.705 and 1.686 species occur on all chromosomes, but that the 1.672 satellite is restricted to chromosome 2.

Highly repeated DNA species	Polytene chromosomes from salivary glands				Tissue culture mitotic chromosomes
1.672					Centromeres of all chromosomes — one major autosome particularly
1.686					Centromeres of all chromosomes one major autosome particularly
1.705					Centromeres of all chromosomes — less on X

Figure 2. Summary of cytological hybridization results. Diagram of results of cytological hybridization of ³H-RNAs complementary to the 1.705, 1.686 and 1.672 satellites. The hatched areas indicate labelling. The methods used for salivary gland chromosomes are given in Figure 1, and the methods for mitotic chromosomes in Figure 4.

Although the 1.672 satellite localized to only chromosome 2 in preparations from both male and female larvae, it is possible that it also occurs on the Y chromosome since this chromosome cannot be identified in the chromocentre (Nicoletti and Lindsley, 1960). Blumenfeld and Forrest (1971) have suggested that this satellite does occur on the Y chromosome; they found a correlation between the amount of this satellite and the number of Y chromosomes. We have data consistent with their findings. DNA, isolated from brains dissected from X/O larvae, showed a decreased amount of 1.672 compared with the amount found in X/Y larval brains. Figure 3 shows the comparative recoveries of the satellites in CsCl, actinomycin-CsCl and netropsin-CsCl for DNA from X/O and X/Y brains, and from tissue culture cells. The tissue culture cells (line 2, Schneider, 1972) contained no Y chromosome and their small yield of the 1.672 satellite provides additional support for a Y chromosome location

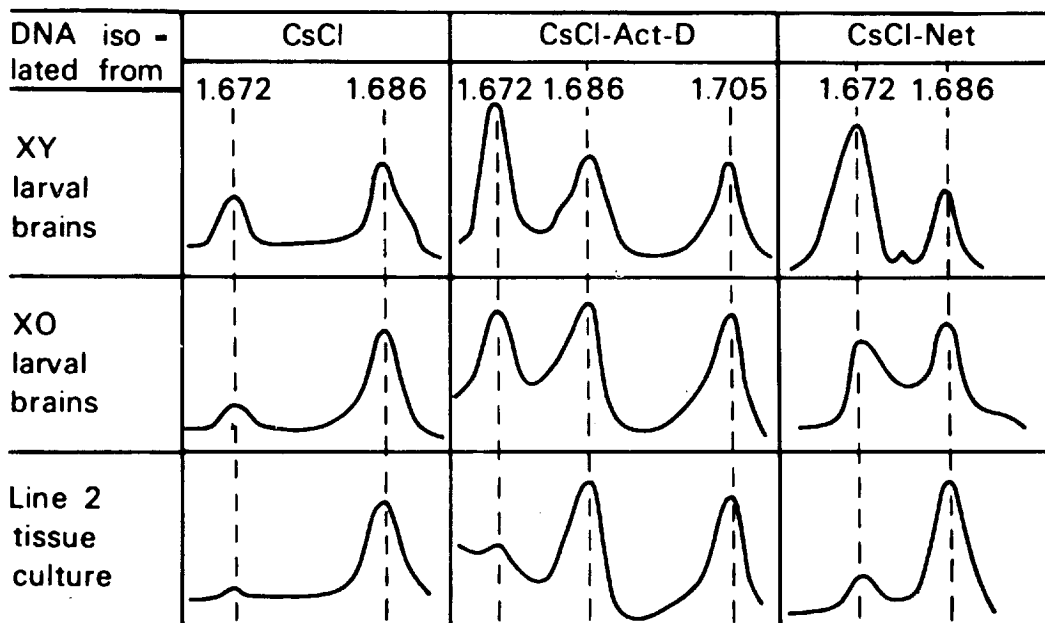


Figure 3. Relative amounts of the satellites present in DNA extracted from X/O and X/Y larval brains and from tissue culture cells. Brains were dissected from larvae in *Drosophila* ringers buffer (Ephrussi-Beadle) and frozen on dry ice. They were lysed in 0.01 M Tris, 1 mM EDTA, pH 8.4, with 1% sarkosyl, and CsCl added to a density of 1.70 g/cc. The lysates were centrifuged for 48 hr at 40,000 rpm. The DNA was then run in an analytical CsCl gradient. Tissue culture cells (line 2, Schneider, 1972) were harvested and lysed with 1% sarkosyl in 0.01 M Tris, 1 mM EDTA, pH 8.4 and processed as above. Actinomycin and netropsin gradients were as described previously (Peacock et al., 1974).

of this highly repeated species. Cytological hybridization to tissue culture metaphase cells (lines 1 and 2, Schneider, 1972) has confirmed the centromeric location of all of the satellites. However these cells, whilst showing the 1.672 species to be predominantly on chromosome 2 and the presumptive Y (in line 1), also showed that it may occur, to a small extent, on chromosome 3 (Figure 4). The small amount of labelling on chromosome 3 may result from other satellites contaminating the 1.672 template, but this is unlikely. If the 1.672 satellite is restricted to chromosomes 2 and Y, *in vivo*, the possibility arises that the two related sequences of this highly repeated species (Brutlag and Peacock, 1975) are distributed with chromosomal specificity. It will be important to isolate this satellite separately from males and females to determine whether both of the sequence types occur in each sex. The tissue culture cells show the 1.686 and 1.705 satellites to be present on all chromosomes, in differing proportions, with the 1.705 occurring to a limited degree in the X chromosome. The 1.686 species appears to be largely confined to one of the major autosomes.

These results, obtained with tissue culture cells, will have to be confirmed with hybridization to *in vivo* mitotic cells since the culture lines do not have the standard male and female chromosome complement. Although the chromosomal morphologies are similar to the normal chromosomes, there could have been rearrangements affecting the amounts and distribution of heterochromatin.

ARRANGEMENT OF THE SATELLITE SEGMENTS

Since cytological hybridization has shown each chromosome of the complement to contain two or more species of satellite in its centromeric region we have attempted to determine the arrangement of the satellites with regard to each other, and with respect to the unique and less repeated DNA of the genome. We have examined the 1.705 satellite and attempted to define its neighbouring sequences; 1.705 DNA from an actinomycin-CsCl gradient was rerun in a CsCl gradient following extraction of the actinomycin. This DNA banded in a single peak but showed a pronounced skew on the light side, suggesting that other DNA species were covalently attached to the 1.705 DNA (Figure 5). The peak was divided into three fractions and each was analysed by buoyant density centrifugation, thermal denaturation and by hybridization. The buoyant density profiles (Figure 5) show fraction 1 to be a single symmetrical band of density 1.705, fraction 2 to have a single band with some skew to the light side, and the band in fraction 3 to have a pronounced shoulder at a density of 1.697 g/cc. Since each of the 1.672, 1.686 and 1.705 satellites has a distinctive T_m , different from that of main band DNA, we were able to use thermal denaturation analysis to monitor the sequence composition of the three fractions. Fraction 1 showed a monophasic

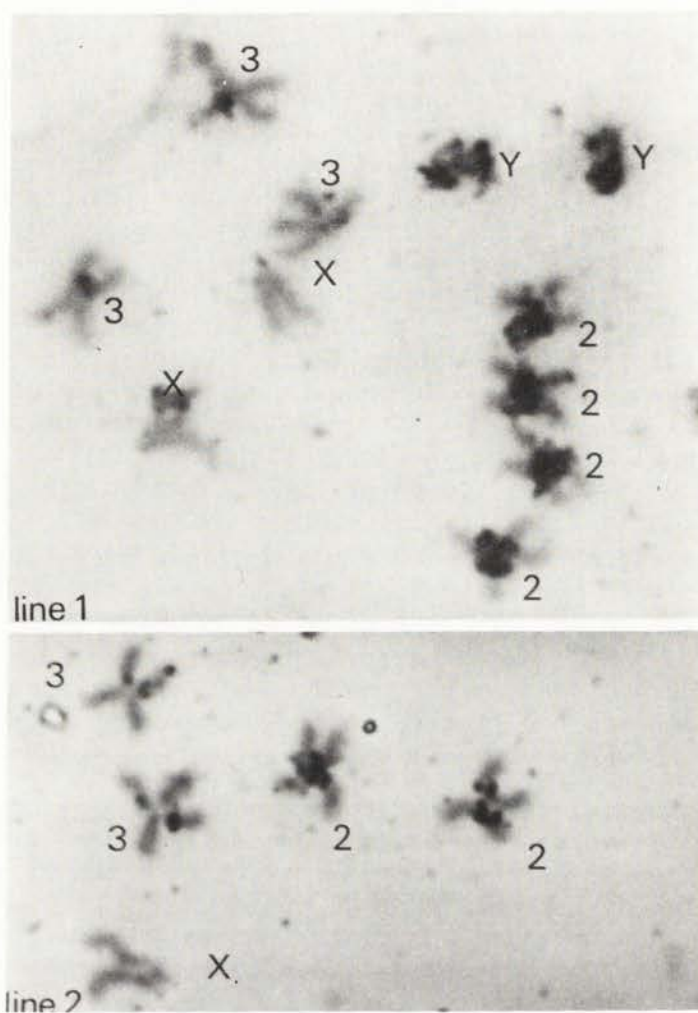


Figure 4. Cytological hybridization of ^3H -RNA, complementary to the 1.672 satellite, to mitotic chromosomes from tissue culture cells. Chromosomes were prepared by the addition of colchicine (2.5×10^{-6} M) to early log phase cultures (lines 1 and 2, Schneider, 1972) for 16 hr. Cells were harvested and resuspended in 0.075 M KCl for 4 min. Cells were fixed, 5 to 30 min, in 3:1 ethanol acetic acid, washed in fresh fixative and resuspended in a small amount of fixative and dropped on to a slide. The chromosomes were stained with 2% orcein for 2 hr prior to cytological hybridization procedures (see Figure 1).

melt with a T_m identical to that of pure 1.705. Fraction 2 showed a polyphasic melt with transitions corresponding to the T_m 's of 1.686 and main band DNA as well as 1.705. Fraction 3 showed these same transitions with an increased amount of main band, but also

showed a small amount of the 1.672 satellite. We were not able to monitor the presence of the 1.688 satellite because its T_m is indistinguishable from that of the 1.705 species.

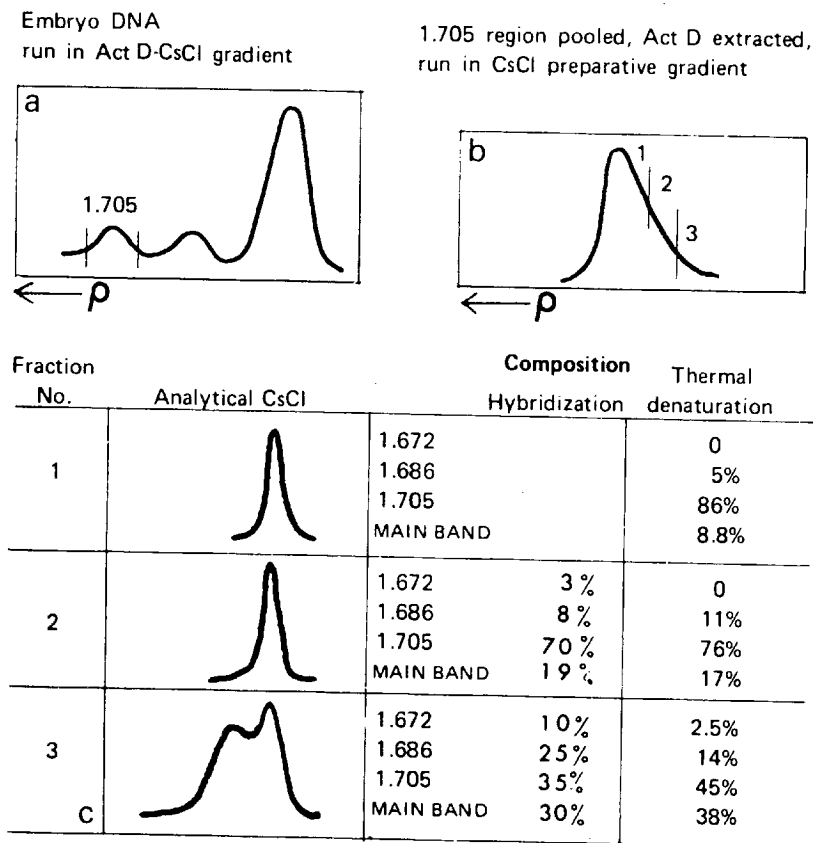


Figure 5. Analyses of DNA species present at the 'ends' of 1.705 molecules. Nuclear DNA from 8 hr embryos was run in actinomycin-CsCl. (a) The 1.705 peak was collected, actinomycin extracted and a second CsCl preparative gradient run. (b) The 1.705 peak was divided into 3 regions as indicated and each fraction analyzed by analytical centrifugation, thermal denaturation and filter hybridization. (c) Thermal denaturation was by the method of Mandel and Marmur (1968). In the hybridization analysis each of the three fractions was fixed to a filter by the method of Gillespie (1968). Small filters were cut and pre-incubated in 0.02% polyvinyl pyrrolidone and 0.02% ficoll. Radioactive DNA was prepared by copying purified species with DNA polymerase in the presence of ^3H -triphosphates. DNA (40,000 counts/min/ μg approximate specific activity) was denatured by heating to 90°C for 2 min. The filters were added and incubated in $2 \times \text{SSC}$, with shaking, carried out at $T_m - 25^\circ$ for each of the satellite species over a period of 12 hr. For the main band hybridization, sufficient DNA was added so that hybridization could be carried out to a $\text{Cot}_{1/2} = 1,000$. Filters were then washed three times in $2 \times \text{SSC}$, dried and counted.

Filter hybridization analysis, using radioactive DNA copies of each of the satellites and of main band, confirmed the thermal denaturation results. The estimate of covalently attached main band sequences could be inflated by a contaminating tail of the very large main band peak in the initial actinomycin gradient, but the 1.686 estimate is free from this error as the second CsCl gradient would have removed any non-covalently attached sequences of this satellite. Covalently attached 1.686 sequences, in molecules which contain the junctions of 1.705 and 1.686 segments, account for approximately 2% of the DNA in the 1.705 peak. Since the average length of the molecules in the gradient is about 15 kb, the junction molecules will contain, on average, 7.5 kb of 1.686. We have estimated the total amount of 1.705 DNA in the genome to be 6,000 kb, so the 2% contamination by 1.686 is consistent with there being a limited number of 1.705-1.686 junctions and with there being a limited number of long segments of the 1.705 sequence. The analysis has shown that the 1.705-1.672 junctions are only one third as frequent as 1.705-1.686 junctions, which is consistent with the 1.672 satellite being restricted to two chromosomes.

DISCUSSION

Our experiments show that the major satellites of *Drosophila melanogaster* occur in long segments at only a few sites in the genome. The shearing-yield experiments have led to estimates of c.600 kb for the length of uninterrupted stretches of each of the highly repeated DNA species. These estimates contrast with those of Kram et al. (1972) who concluded that the rapidly renaturing sequences of *Drosophila* occur in stretches of only 2 to 18 kb, interspersed with segments of more complex DNA. In large part their failure to detect long stretches of satellite can be attributed to their method of DNA preparation which results in the loss of most of the satellites (Peacock et al., 1974). The relation of their hydroxylapatite rapidly renaturing (h.a.r.r.) DNA to our nuclear satellites is uncertain. Junctions between satellites and main band DNA could yield molecules corresponding to the h.a.r.r. DNA. We do have some evidence for junctions between the 1.705 satellite and main band sequences but quantitatively these involve only a few per cent of the satellite sequences.

There has been another report of satellite sequences occurring in long stretches. Bond et al. (1967) concluded for the mouse genome that 'nearly all of the satellite sequences do indeed occur in such stretches (60 kb)'. It is likely that in all species in which cytological hybridization has shown a centromeric location of highly repeated DNA, that these sequences occur in long segments.

We pointed out earlier that the restriction of the 1.672 satellite to two chromosomes could mean that each one of its two repeating sequences (Brutlag and Peacock, 1975) occurs on only one chromosome. However our cytological hybridization data rule out the possibility that each of the major satellite species occurs on only one chromosome. Nevertheless the *Drosophila* satellites could be distributed in a pattern which is chromosome specific. Both arrangement and length of segments of the different repeating species could distinguish the centromeric heterochromatin of a chromosome from that of all other chromosomes in the complement. Findings of individual satellites occurring on only some members of a chromosome complement (Jones et al., 1973; Jones and Corneo, 1971; Prescott et al., 1973) reflect some chromosomal specificity of highly repeated DNA species. It should be possible in *Drosophila*, by manipulating the chromosome complement, to determine whether a chromosome specific arrangement of the satellites does exist.

A surprising result of the cytological hybridization experiments was the finding that each of the three satellites (1.672, 1.686 and 1.705) hybridized to region 21C-D on chromosome 2. The 1.705 localization to 21C-D has been checked by competition hybridization analysis to be sequence specific and this undoubtedly applies to the other two satellites. We have calculated from the relative number of grains over 21C-D and over the chromocentre that the 1.705 sequences at the 21C-D region account for only 6 kb of the total of 6,000 kb of 1.705 DNA in the genome (Peacock et al., 1974). The calculation is dependent on a number of assumptions; however, the fact remains there is a small representation of each of the three satellites in what may be a single band on chromosome 2. The explanation for this may be trivial; it may be a consequence of two or more inversions which involved break points in the heterochromatin resulting in the transposition of portions of the heterochromatin to 21C-D. On the other hand this locus may have a fundamental role in the establishment and maintenance of satellite sequences. It will be of some interest to examine other *Drosophila* species to determine whether there is a similar band in the complement, whether it is in a homologous position in the complement, and whether it contains each of the centromeric satellite sequences.

We have previously pointed out a correspondence in amount and location of heterochromatin and the satellite sequences (Peacock et al., 1974). However, this may be an oversimplification. The tissue culture chromosomes have shown satellite localization only to the centromeric heterochromatin of the X chromosome and not more generally over the basal heterochromatin of this chromosome. Thus the satellite species may be restricted to block A of the heterochromatin (Cooper,

1959). The nucleolus organizer region, the site of the ribosomal RNA genes, occurs between blocks B and C, yet filter hybridization with radioactive ribosomal RNA has given no evidence for ribosomal cistrons in any of the three purified satellite species. This is consistent with them having a restricted distribution in the basal heterochromatin. We did find an enrichment of ribosomal sequences in the preparation of 1.688 DNA. However, until we have further purified this satellite the significance of this finding is uncertain and may merely represent a co-purification of 1.688 and ribosomal DNA in the actinomycin gradient.

If the satellites are restricted to block A of the X chromosome heterochromatin this may conflict with our suggestion of chromosomal specificity of their arrangement. Chromosomal specificity implies an importance in homologous chromosome recognition, yet Cooper (1964) has shown that block A lacks sites for meiotic X and Y chromosome conjunction in the male. Nevertheless, even if highly repeated DNA does not play this role in meiosis in the male, it may in meiosis in the female, or in other processes in which homologous chromosome recognition is important.

REFERENCES

- Blumenfeld, M. and H.S. Forrest. 1971. Is *Drosophila* dAT on the Y chromosome? *Proc. Nat. Acad. Sci. USA* 70: 535.
- Bond, H.E., W.G. Flamm, H.E. Burr and S.B. Bond. 1967. Mouse satellite DNA : Further studies on its biological and physical characteristics and its intracellular location. *J. Mol. Biol.* 27: 289.
- Brutlag, D.L. and W.J. Peacock. 1975. Sequences of highly repeated DNA of *Drosophila melanogaster*. This volume.
- Cooper, K.W. 1959. Cytogenetic analysis of major heterochromatic elements (especially Xh and Y) in *Drosophila melanogaster*, and the theory of heterochromatin. *Chromosoma* 10: 535.
- _____. 1964. Meiotic conjunctive elements not involving chiasmata. *Proc. Nat. Acad. Sci. USA* 52: 1248.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology. In *Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 21D, p.413. Academic Press, New York.
- Gillespie, D. 1968. The formation and detection of DNA-RNA hybrids. In *Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 12B p.641. Academic Press, New York.
- Jones, K.W. and G. Corneo. 1971. Location of satellite and homogeneous DNA sequences on human chromosomes. *Nature New Biol.* 233: 268.

- Jones, K.W., J. Prosser, G. Corneo and E. Ginelli. 1973. The chromosomal location of human satellite DNA III. *Chromosoma* 42: 445.
- Kram, R., M. Botchan and J.E. Hearst. 1972. Arrangement of the highly reiterated DNA sequences in the centric heterochromatin of *Drosophila melanogaster*. Evidence for interspersed spacer DNA. *J. Mol. Biol.* 64: 103.
- Mandel, M. and J. Marmur. 1968. Use of ultraviolet absorbance - temperature profile for determining the guanine plus cytosine content of DNA. In *Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 12B p.195. Academic Press, New York.
- Nicoletti, B. and D.L. Lindsley. 1960. Translocations between the X and Y chromosomes of *Drosophila melanogaster*. *Genetics* 45: 1705.
- Pardue, M.L. and J.G. Gall. 1970. Chromosomal localization of mouse satellite DNA. *Science* 168: 1356.
- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 405.
- Prescott, D.M., C.J. Bostock, F.T. Hatch and J.A. Mazrimas. 1973. Location of satellite DNAs in the chromosomes of the kangaroo rat (*Dipodomys ordii*). *Chromosoma* 42: 205.
- Schneider, I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morph.* 27: 353.

The organization of eukaryotic DNA and the sequence environment of structural genes

Eric H. Davidson, Barbara R. Hough,
Michael J. Smith, Dale E. Graham,
William H. Klein, Glenn A. Galau,
Margaret E. Chamberlin, and Roy J.
Britten

Recent, detailed studies of the organizational characteristics of eukaryotic DNA are reviewed in relationship to the implications of the physical patterns of DNA sequence arrangement for an understanding of the mechanism of control of gene activity. The interspersion of 300 nucleotide long repetitive DNA sequences with nonrepetitive sequences 1,000 to more than 4,000 nucleotides in length, previously shown in reassociation experiments with the DNAs of *Xenopus* and of sea urchin, is confirmed by measurements of electron micrographs of these DNAs. Structural genes of the sea urchin, identified by hybridization with messenger RNA released from polysomes, are characterized as representing in the main nonrepetitive sequences, of a length similar to the interspersed nonrepetitive DNA lengths. They include tens to hundreds of thousands of nucleotide pairs of sequence complexity, and are located adjacent to repetitive sequences in the genome.

Co-ordination of the activities of large numbers of genes is a basic characteristic of differentiated eukaryotic cells. To understand the physical basis for this aspect of gene regulation it seems evident that specific knowledge of the patterns of DNA sequence organization will be needed. In particular one is concerned with aspects of genomic organization which might relate different structural genes in the overlapping functional assemblages operating in living cells. For instance, one possibility first proposed by us several years ago is that functionally related sets of structural genes share a common repetitive sequence (Britten and Davidson, 1969). Whether or not this model is correct, the path of reasoning to which it has already led suggests that the properties of gene regulation systems in animals probably derive from the particulars of genomic sequence organization.

Recently our laboratory has been engaged in a detailed study of the organizational characteristics of eukaryotic DNA. We have discovered that the animal genome is indeed a highly ordered structure. Most of the DNA in two very different creatures, the amphibian *Xenopus laevis* and the sea urchin *Strongylocentrotus purpuratus*, has been found to consist of short, repetitive sequence elements only a few hundred nucleotides long, interspersed among nonrepetitive sequences of one to several thousand nucleotides. In the present

paper we report further independent evidence regarding quantitative aspects of sequence organization in *Xenopus*. In addition we summarize the status of evidence regarding sequence interspersion in a variety of other DNAs. We then describe briefly some recent experiments which begin to penetrate directly to the issue of sequence organization in the immediate environment of working structural genes.

SEQUENCE ORGANIZATION IN ANIMAL DNAs

Status of knowledge regarding sequence organization in *Xenopus* and sea urchin DNAs Since space is limited, we cannot recapitulate in detail information of DNA sequence arrangement which is already published elsewhere. A brief summary of our present knowledge of *Xenopus* and *Strongylocentrotus* DNAs now follows.

(1) Over 80% of the DNA contains interspersed repetitive sequence elements. We know this from studies of hydroxyapatite binding of long segments of labeled DNA which has been reassociated with excess short DNA fragments to a C_{0t} where only repetitive sequences react (C_{0t} , concentration \times time, is expressed in terms of moles nucleotide litre⁻¹ \times sec). The mixture is passed over a hydroxyapatite column and those fragments bind which contain a repetitive sequence duplex. Such experiments were carried out by Britten (1972) who used a particular repetitive fraction of sea urchin DNA and a single tracer fragment length. The approach was then extended by the use of various measured lengths of labeled DNA to *Xenopus* (Davidson et al., 1973) and to sea urchin (Graham et al., 1974).

(2) The interspersed repetitive sequence regions are separated by nonrepetitive DNA sequences ranging in length from about 1,000 to >4,000 nucleotides. Quantitative estimates of inter-repeat spacings are derived directly from the measurements of hydroxyapatite binding of labeled DNA fragments as a function of tracer fragment length (Davidson et al., 1973; Graham et al., 1974). About half of the single-copy sequences are $1,000 \pm 300$ nucleotides, and the remainder are significantly longer.

(3) About 25% of the total DNA nucleotides are in repetitive sequences, and 75% are in nonrepetitive sequences. This fact can also be derived from the hydroxyapatite binding fragment length experiments. Another way in which the same result is obtained is measurement of the amount of hyperchromicity developed when fragments of different lengths containing repetitive sequence duplexes are melted in a spectrophotometer cuvette (Davidson et al., 1973; Graham et al., 1974). A third approach, completely independent of the last two, involves the application of 'S-1 nuclease', a single-strand specific endonuclease. The nuclease is used to isolate repetitive duplexes

by digesting away single-stranded regions. The amount of enzyme resistant duplex can then be quantified (Britten, Graham, Davidson and Eden, in preparation).

(4) About three-quarters of the repetitive sequences exist in elements which are only 300 ± 150 nucleotides in length. This estimate comes most directly from the S-1 nuclease studies just cited, where the size of the enzyme-resistant duplex regions is measured by gel filtration. Corroborative information comes from experiments in which the fraction of long tracer fragments bearing repetitive sequence elements actually in the repetitive sequence(s) is measured by studying their rate of reassociation with repetitive DNA (Davidson et al., 1973). An average length of 300 nucleotides also derives from the quantitative evaluation of hyperchromicity, mentioned above (Graham et al., 1974; Davidson et al., 1973). The S-1 nuclease experiments identify the presence of some long repetitive sequence regions (i.e. $>1,000$ nucleotides) as well. So far we know relatively little about these. About one-quarter of the repetitive sequence, or 6-8% of the whole DNA, is in long repetitive regions.

(5) Short (300 nucleotides) interspersed repetitive sequence regions characteristically display about 7-15% sequence divergence. This conclusion is drawn from analysis of the thermal stability, after S-1 nuclease treatment to remove unhybridized single stranded tails (Britten et al., in preparation). The effects on hyperchromicity and thermal stability of sequence mismatch within a repetitive duplex and the effects of totally single-stranded tails can now be distinguished. Interestingly, the same thermal stability experiments reveal that the long repetitive sequences display a relative lack of sequence mismatch. This observation clearly poses some fascinating problems in DNA evolution.

We conclude that the DNAs of *Xenopus* and sea urchin are generally ordered in a highly specific way. The key feature is the very extensive interspersion of repetitive sequence elements which appear too short to code for most known proteins (average 300 nucleotides). Furthermore, the interspersed repetitive sequences exist in many mutually non-homologous species. For example, in *Xenopus* the best estimates are that there are about 2,000 diverse species of repetitive sequence, each present 1,500 to 2,000 times per genome (Davidson et al., 1973). Thus, the interspersed repeats constitute a relatively complex group of sequences completely distinct from the extremely simple satellite DNAs discussed elsewhere in this volume.

Sequence interspersion viewed in the electron microscope Four essentially separate and independent methods of analysis have led to

the knowledge summarized above (see Table 2, section 2d). In this section we describe some results obtained with a fifth approach, direct observation with the electron microscope. The technique we have employed is the formamide-Kleinschmidt method developed by Davis et al. (1971), in which single-strand as well as double-strand regions can be visualized. The main object of these experiments is to obtain independent corroboration for the length of 300 nucleotides (av.) which we have assigned to interspersed repetitive sequences on the basis of the S-1 nuclease and other studies.

Our evidence leads us to believe that most repetitive sequence elements are flanked by single-copy regions, one to several thousand nucleotides long. It follows that duplexes formed from fragments of such lengths should be present in the form of four-ended structures which should be easily visualized in the electron microscope. That is, we would expect to see in the electron microscope a large number of short (av. 300 nucleotides) duplex regions, each end of which gives rise to two longer single-stranded tails. The interspersion data predicts that a fraction of longer fragments would contain two or more repetitive sequence regions, each of which should be marked by the presence of similar single-stranded tails.

Examples of both the simple four-ended structures and more complex molecules displaying more than one interspersed repetitive sequence are shown in Figure 1. The *Xenopus* DNA portrayed in this figure had been sheared to a modal length of about 2,000 nucleotides (though a broad range of sizes is of course present), and then re-associated to C_0t 10. By this point almost all the repetitive sequences in *Xenopus* DNA will have formed duplexes. The duplex-containing fraction was recovered by passage over a hydroxyapatite column, and then spread for electron microscopy as described in the legend to Figure 1. This series of experiments is treated in detail elsewhere (Chamberlin, Britten and Davidson, in preparation), and only a brief account of some of the main points is possible here.

The structures shown in Figure 1 represent major classes in the preparations. This is a key point, for it is very easy to be misled in molecular electron microscopy unless the total mass of DNA present can be accounted for. Table 1 provides a complete breakdown of the 1,660 structures present in the EM plates we studied. It can be seen that four-ended structures in which the terminations of the duplex regions are both unequivocally marked by single-strand forks represent 28% of the DNA. Molecules displaying more than one interspersed repetitive region ('complex structures' in Table 1) account for another 24%. Thus over half the DNA is present in molecules containing *duplexes marked by single-strand tails on both ends of each duplex*. Only 5% of the DNA mass is in uninterpretable structures

TABLE 1. Electron micrograph analysis of *Xenopus* DNA reassociation products after incubation to low C_{ot}^a

Class of structures	Number scored	% of total structures	% of structures containing interpretable duplexes	Total strand length (nucleotides)	% length in class	Length of molecules (nucleotides)	Mean	Mode class
Four-ended structures	362	21.8	52.62	0.88×10^6	28.03	2445	2445	2400
Three-ended structures	40	2.41	7.27	0.12×10^6	3.82	3022	3022	2700
Complex molecules ^b	276	16.63	40.11	0.77×10^6	24.52	2774	2774	2700
No duplex regions with tails, including possible two-ended duplexes	932	56.15	-	1.2×10^6	38.22	1311	1311	1200
Uninterpretable structures	50	3.01	-	0.17×10^6	5.41	3429	3429	3000
Total	1,660			3.14×10^6				

^a *Xenopus* DNA, prepared as described in the legend to Figure 1, was incubated to Cot 10, at which repetitive sequences form duplexes but nonrepetitive sequences do not. The fraction which bound to hydroxyapatite was eluted and prepared for electron microscope visualization. Chamberlin, Britten and Davidson, in preparation.

^b This classification consists mainly of molecules bearing more than one duplex region. Also included are multibranch structures, circles and loops (5 in all). All of these molecules contained interpretable duplex regions.

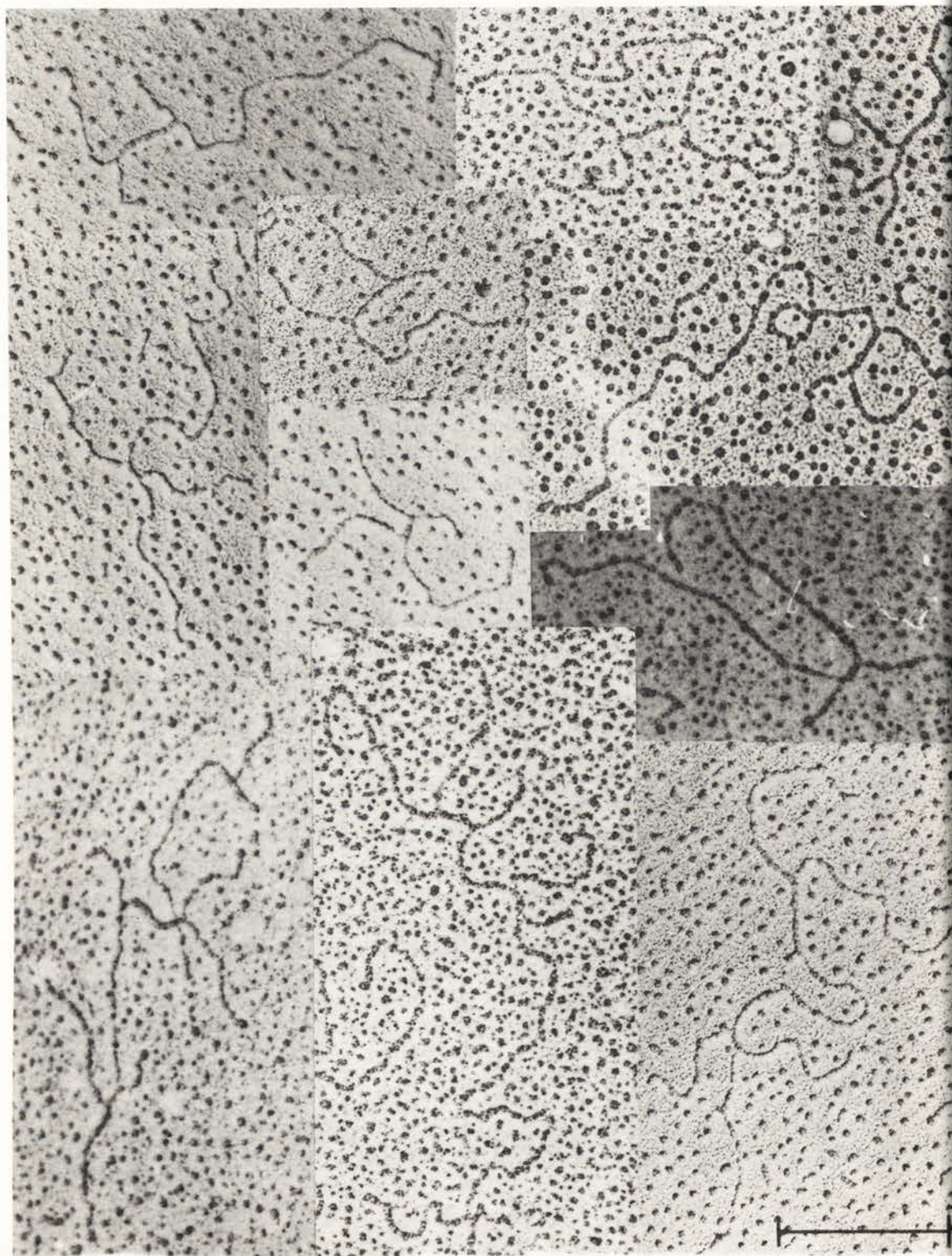


Figure 1. Direct visualization of interspersed repetitive duplex regions in *Xenopus* DNA (Chamberlin, Britten and Davidson, in preparation). *Xenopus* DNA was sheared to a modal size of 2,400 nucleotide pairs in a Virtis homogenizer at 6,000 rpm for 45 min. The DNA was dissolved in 0.12 M phosphate buffer and was kept at 0°C during shearing. The DNA was denatured at 100°C for 5 min and immediately passed over a hydroxyapatite column in 0.12 M phosphate buffer at 60°C to remove the 'fold-back' fraction (Davidson et al., 1973; Wilson and Thomas, 1973). 30% of the DNA bound to the column. The non-binding fraction was found to be the same size as the starting material. It was incubated to C_{0t} 10 and passed over a hydroxyapatite column. 56.6% of the DNA bound to the column. The bound fraction was eluted with 0.5 M phosphate buffer, dialyzed against 50% formamide-0.01 M EDTA-0.1 M Tris, pH 8.5, and spread for electron microscope visualization from 57% formamide (Davis et al., 1971) by the Kleinschmidt procedure. The structures shown in the eight upper photographs each contain a duplex region and four single-stranded ends. The three lower photographs represent molecules which include more than one duplex region, each marked by single-strand tails (Chamberlin, Britten and Davidson, in preparation). The bar represents a length of 1,000 nucleotides.

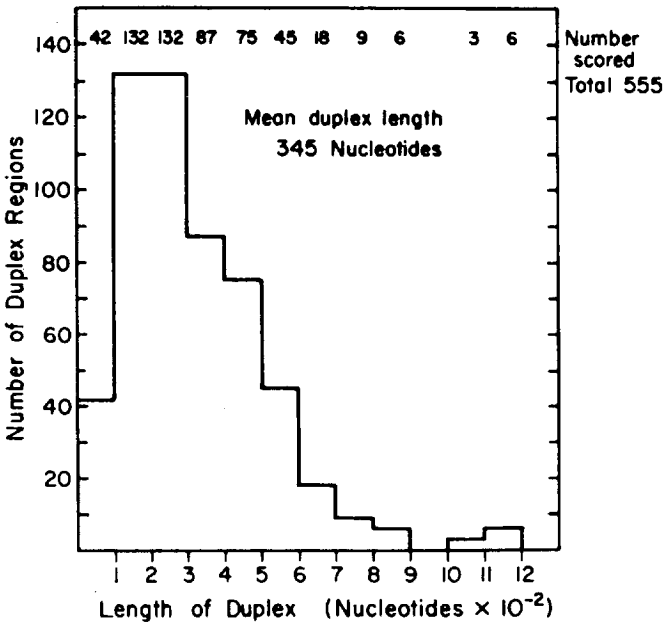


Figure 2. Repetitive sequence length in *Xenopus* DNA. Length in nucleotides of duplex regions was measured from electron micrographs such as those shown in Figure 1. All duplexes on both four-ended structures and more complex interspersed molecules are included in this analysis, except that the histogram excludes length classes for which less than three examples were measured. The class width was 100 nucleotides. The number of duplexes falling in each class is given on the top of the figure. The longest clearly evident duplex measured contained 2,850 nucleotides; the shortest, about 60. The mean duplex length for the sample was 345 nucleotide pairs.

The fragments present in all of the classes of structure containing evident duplex are about the same length (mean 2,500-3,000 nucleotides), but this is not true for the two-ended structure class. The latter should include any long duplexes from which no tails or only one single-strand tail protrudes, but it may also include all broken single-strand ends originally part of duplex-containing molecules. The distinctively smaller size of the two-ended fragments indicates that a large fraction of this class (mean length 1,311 nucleotides) does consist of broken ends. As proof of this point we found that 70-100% of fragments in each length class over 2,000 nucleotides contain recognizable duplexes, while only 15% of fragments less than 1,200 nucleotides contain recognizable duplexes.

The four-ended structures and complex molecules are excellent for direct estimation of repetitive duplex length, since for each molecule in this class the ends of the duplex are clearly marked by single-stranded tails (see Figure 1). A histogram representing the repetitive duplex length distribution is shown in Figure 2. Though the distribution is skewed toward longer lengths, indicating the presence of a few long duplex regions, the main result of this analysis is clear. The interspersed repetitive sequence elements are only a few hundred nucleotides in length. In fact, 60-70% of all duplexes seen are 300 ± 150 nucleotides. Furthermore, this statement remains true for all fragment lengths between 1,000 and 6,000-8,000 nucleotides, where our observations end. It follows that the sample from which duplex length is estimated has not been biased by fragment length. That is, we are not losing longer duplexes from the sample by working at the fragment lengths we have chosen, since if that were the case an increase in mean duplex length with fragment length would result.

Those structures which permit measurement of inter-duplex length (e.g. the three lower molecules in Figure 1) are the subject of a further analysis shown in Figure 3. The distribution of inter-duplex single-copy lengths measured in this sample can be used to calculate the expected relation between hydroxyapatite binding and fragment length. The calculated estimate is compared in Figure 3 with the binding curve measured in the experiments summarized above. The conclusion from this analysis is that the repeat spacing distribution observed in the electron microscope data is consistent with the spacing distances we have deduced from the hydroxyapatite binding experiments, within our limits of accuracy.

To summarize this brief account, these new studies of *Xenopus* DNA reassociation products have again shown, this time by direct observation in the electron microscope, that repetitive sequence elements are interspersed with nonrepetitive sequences. Furthermore,

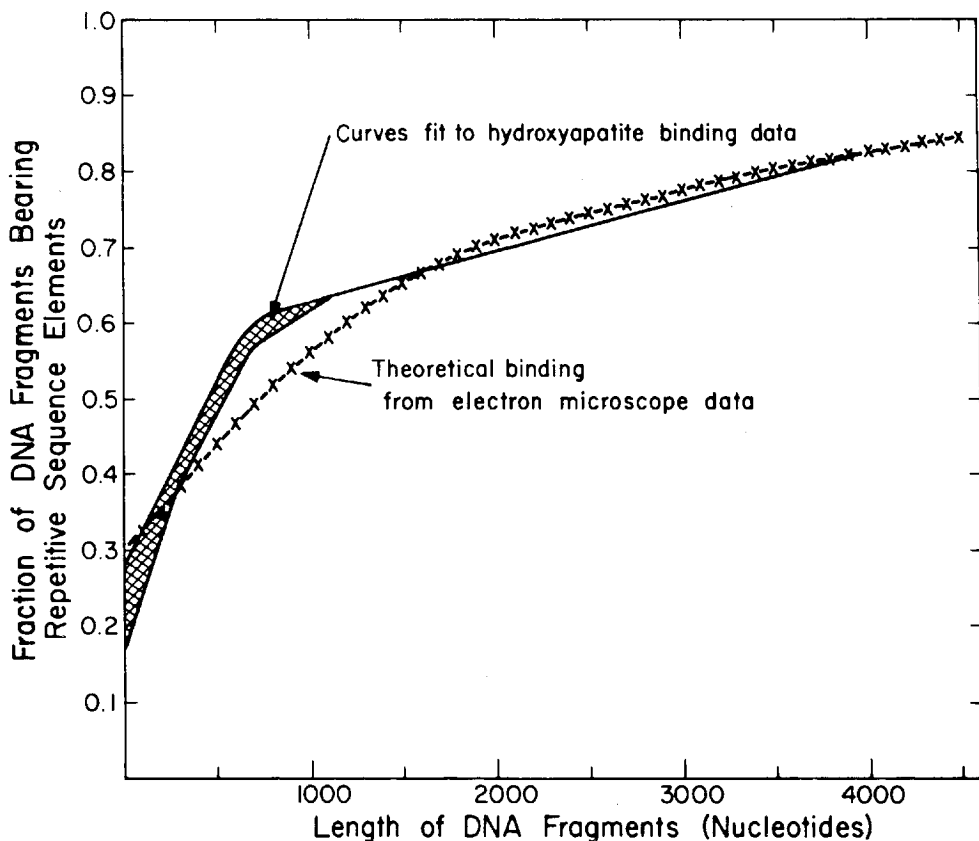


Figure 3. Comparison of calculated (-X-) and observed (—) relation between fragment length and presence of a repetitive sequence element on the fragment (Chamberlin, Britten, and Davidson, in preparation). The observed curve is the corrected hydroxyapatite binding curve reported by Davidson et al. (1973) from *Xenopus* DNA (see text). The hatched area represents the range of acceptable fits presented in this reference. The calculated curve is derived from the distribution of nonrepetitive sequence lengths measured in molecules where such sequences could be defined as bounded on both ends by repetitive sequence elements. Examples are shown in the bottom three structures of Figure 1. To avoid biasing the sample in the direction of shorter nonrepetitive sequence lengths, only molecules longer than 3,000 nucleotides were included. The number of available cases was 77, and the fragment lengths were, of course, insufficient to provide accurate data for the terminal portions of the curve. Therefore, in this calculation the outer slope of the curve (fragment lengths >3000 nucleotides) was matched to that of the hydroxyapatite curve. The remainder of the curve was calculated from the interduplex length distribution on the basis that the rate of change in slope (second derivative) of the curve is related to the frequency with which interduplex single-copy sequences of given lengths occur.

the mean length of most of these sequences is found to be close to 300 nucleotides, in accord with the conclusions drawn from physical chemical data. Our earlier conclusions on the distribution of non-repetitive sequence lengths are also supported by the available electron microscope measurements.

Homologous repetitive sequence elements have homologous terminations

An interesting question, and one which is of fundamental importance in considering possible functions of repetitive DNA, is whether a given interspersed repetitive sequence begins and terminates at the same place in the sequence wherever it occurs. One can imagine, though it would be something of a conceptual nightmare, that a repetitive sequence a-z would begin at n and terminate at m at one site, begin at s and terminate at r in another, and begin at a and terminate at z in still another, etc. The alternative is that each time this sequence appears it begins at a and ends at z. A simple way to determine which model more nearly approximates the case has been devised, and we present an example of a type of experiment capable of deciding this question. A more extensive description of these experiments will shortly appear elsewhere (Graham, Eden, Davidson and Britten, in preparation).

We begin by considering DNA containing no sequence repetition, e.g., a prokaryote DNA. If this DNA is sheared randomly and allowed to reassociate, very large particles ('hyperpolymers') will eventually form. The explanation for hyperpolymer formation is that on the average no two sheared fragments which meet in the solution, collide and nucleate to form duplex, begin and end at the same place in the sequence. Hence every collision and subsequent duplex formation tends to leave one or more unpaired single-stranded tail. Each of these tails can next participate in a further renaturation reaction with another single-strand region, which as the reaction proceeds, is itself increasingly likely to belong to another growing hyperpolymer. In this way large multimeric hyperpolymers will build up, and late in the reaction these become the dominant type of structure present.

With this background we now consider the likely outcome of a renaturation reaction between fragments bearing repetitive sequence elements. Several cases can be distinguished. For clarity we first consider a case where the fragment length is much longer than the interspersed repetitive sequence and the homologous repetitive sequence elements always begin and terminate at similar locations. Here the problem is trivial since in general only four-ended structures such as those shown in Figure 1 would be able to form. Four-ended dimers would thus represent the terminal reaction products. Suppose now that the repetitive sequence elements are isolated by

treatment of preparations such as those in Figure 1 with S-1 nuclease, and the enzyme-resistant DNA is denatured and then reassociated. It is obvious that if each type of repetitive sequence element always begins and ends at the same place, hyperpolymer formation will still be severely limited. Even fragments bearing partial sequences will not cause significant hyperpolymer formation. This is because the typical 300 nucleotide sequence cannot be divided up between many participating fragments, each maintaining a duplex region sufficiently long to ensure stability. The most likely structures very late in the reaction are dimers, with some trimers and tetramers also present. On the other hand, suppose the S-1 nuclease-resistant duplexes contain parts of repetitive sequence elements which begin and terminate at various places throughout the sequence. In essence this case is similar to that of the sheared prokaryote DNA. Thus large, branched hyperpolymers would be expected to form during an extensive renaturation carried out with such fragments.

Distinguishing the above cases in terms of the predicted particle size of late reaction products means that the problem can be approached empirically simply by measuring the approximate size of the structures in a renatured preparation. In Figure 4 a representative experiment is portrayed. Here we have used an agarose gel filtration column to separate dimers and low multimers from large hyperpolymers. The panel on the left represents the elution of T-4 phage DNA sheared to 300 nucleotides and renatured to about 80 times the $C_{0t_{1/2}}$. It is clear that most of the DNA is excluded, meaning that as predicted, most of the 300 nucleotide fragments are bound in large hyperpolymers. On the right is shown the elution pattern of repetitive sequence elements isolated from sea urchin DNA with S-1 nuclease, and then renatured to 100 times their $C_{0t_{1/2}}$. Since the resistant DNA fragments are about 300 nucleotides (av.), they are about the same fragment length as the phage DNA on the left. In contrast to the latter, however, almost all the sea urchin repetitive DNA penetrates the agarose column, and chromatographs in the positions expected for particles the size of dimers, trimers, and tetramers. According to the above logic this result provides a clear demonstration that homologous, interspersed repetitive sequence elements tend to begin and to end at the same sequence locations. Such an experiment of course cannot resolve differences in termination position less than the stable duplex length, i.e. about 30-50 nucleotides.

The experiment shown in Figure 4 means that each species of repetitive sequence tends to present the same aspect whenever it occurs in the genome. This is probably a necessary condition for theories of repetitive sequence function which require that the same sequence specific ligand be able to bind to homologous repeats wherever they occur in the genome (e.g. Britten and Davidson, 1969; Davidson and Britten, 1973).

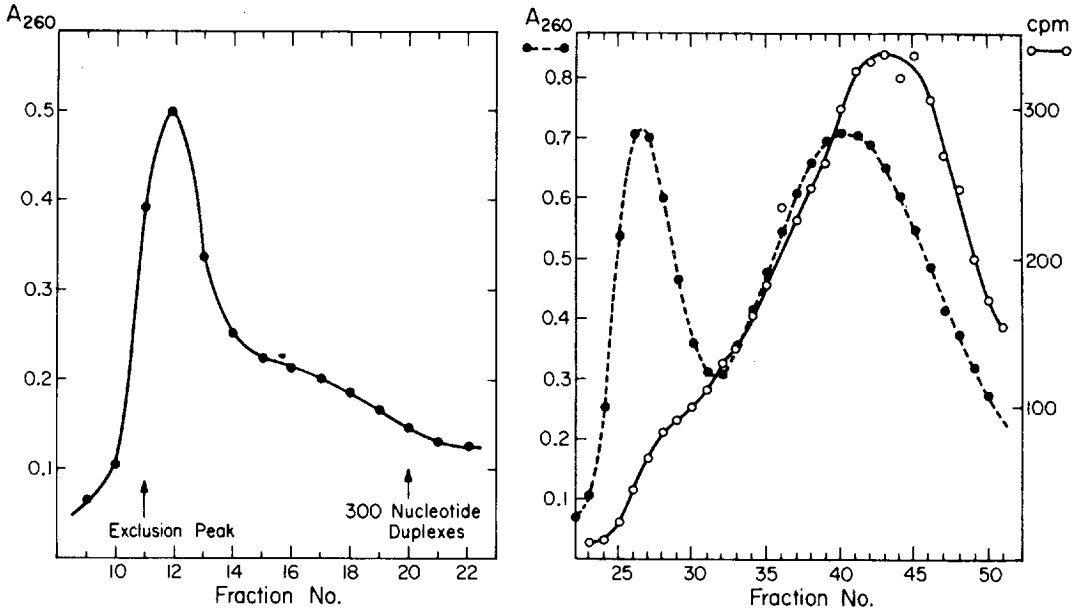


Figure 4. A50 agarose column chromatography of DNA hyperpolymers (Graham, Eden, Davidson and Britten, in preparation). (a) Three hundred nucleotide long fragments of T4 DNA were incubated at 60°C in 0.12 M phosphate buffer to C_{ot} 17.2 (estimated to equal about 80 times the $C_{ot}_{1/2}$ under these conditions). The DNA was placed directly on the column after incubation. 66% of the DNA mass elutes from the column in fractions 10-15, as particles many times larger than 300 nucleotide duplexes. Position of inclusion and exclusion peaks are denoted by arrows. (b) ^3H -labeled repetitive sequences from sea urchin DNA were prepared by reassociation of 2,000 nucleotide long fragments to C_{ot} 1 and subsequent removal of single-stranded material with S-1 nuclease. The interspersed repetitive sequences were isolated, incubated to C_{ot} 11.7 (about 100 x $C_{ot}_{1/2}$) and placed on an A50 agarose column (○-○). Calibration of the column with unsheread calf DNA and 300 nucleotide duplexes is also shown (●- -●).

Sequence interspersion patterns in the genomes of other animals

All of the observations regarding sequence organization so far presented refer to sea urchin and *Xenopus* DNA. These are the DNAs about which most is known and for which widespread sequence interspersion is most clearly and quantitatively demonstrated. However, observations exist for other DNAs as well, and these are summarized in Table 2. As the title of this table indicates, extant data suggest that repetitive sequence interspersion is of general occurrence, perhaps characteristic of all animal DNAs. It is worth noting, however, that for no animals other than *Xenopus* and *Strongylocentrotus* do we have a clear idea of the lengths of single-copy sequence separating adjacent interspersed repetitive sequences. Furthermore, while in both of these genomes the dominant repetitive sequence

element length is about 300 nucleotides and only a minor fraction of the repetitive DNA occurs in longer sequences, this is likely not always to be so. An example may be the *Drosophila* genome. New data of Hogness presented elsewhere in this volume show that at least some middle repetitive sequence interspersions exist in *Drosophila* DNA, but in contrast to other known cases, in *Drosophila* a large fraction of the middle repetitive sequences are apparently significantly longer than a few hundred nucleotides (Manning, Schmid and N. Davidson, in preparation). Clearly much more work is required before quantitative generalizations are feasible.

STRUCTURAL GENE SEQUENCES IN ANIMAL DNAs

In the following discussion we rely on an empirical definition of structural gene sequences, *viz.* those DNA sequences from which functional polysomal mRNA is transcribed. Such sequences can be identified by hybridization with polysomal mRNA. As described elsewhere (Goldberg et al., 1973; Galau et al., 1974), the mRNA used in our experiments is obtained by puromycin release from polyribosomes, and is almost completely free of contamination with hnRNA sequences.

Structural genes are generally single-copy sequences Hybridization experiments have shown that most structural genes belong to the single-copy sequence class. These experiments have been reviewed extensively by Davidson and Britten (1973). Several important conclusions derive from this body of work. A reliable method for measuring multiplicity of particular structural genes is the use of reverse transcriptase to produce a complementary (cDNA) copy of purified mRNA. The cDNA is then hybridized with whole cell DNA and the kinetics analyzed. The cDNA procedure has been used to show that several isolated, purified mRNAs are single-copy transcripts. Such experiments yield easily interpretable kinetic data, since they do not depend on assumptions regarding mRNA-DNA hybridization kinetics, which are found to vary somewhat depending on the RNA used. For example current studies in our laboratory show that while as much as 75% of total sea urchin embryo mRNA can be hybridized at moderate DNA-RNA ratios, the reactions approach termination only at relatively high C_{0t} . That is, the mRNA-DNA duplexes form 2-4 times more slowly than DNA-DNA duplexes.

Among the purified mRNAs studied only histone mRNA hybridizes with repetitive sequences. There may be other yet unidentified structural genes which are present hundreds or thousands of times per genome. Thus, Klein et al. (1974) found that 5-7% of HeLa cell mRNA, exclusive of histone mRNA, hybridizes with HeLa DNA at low C_{0t} . However, the overwhelming majority of mRNAs are transcribed from single-copy sequences. A clear demonstration of this was accomplished by Galau et al. (1974) in an experiment described in

TABLE 2
Evidence for the general occurrence of DNA sequence interspersion

Species	Methods	Conclusions	References
<u>DEUTEROSTOMES</u>			
Amphibians			
<i>Xenopus laevis</i>	A*B*CDE*F	Extensive interspersion: non-repetitive sequences 500 to >4000 nucleotides, short repetitive sequences 300 ± 150 nucleotides	Davidson et al., 1973 Chamberlin, Britten and Davidson, in prep.
Echinoderm			
<i>Strongylocentrotus purpuratus</i>	A*B*E*	Same	Graham et al., 1974
Mammals			
<i>Bos domesticus</i>	ABE*	Extensive interspersion of a repetitive sequence fraction	Britten and Smith, 1970 Authors' unpublished observations
<i>Mus musculus</i>	B	Some short repeated sequences interspersed	Rice, 1971 Britten, 1972
<i>Rattus norvegicus</i>	B	Same	Rice, 1971
<u>PROTOSTOMES</u>			
Gastropods			
<i>Aplysia californica</i>	B*CE*	Most single-copy sequences are interspersed with short repetitive sequences	Angerer, Davidson and Britten, in prep.

PROTOSOMES (contd)

<i>Ilyanassa obsoleta</i>	B	Sequences of at least one repetitive component are short and interspersed	Davidson et al., 1971
Insects			
<i>Oncopeltus fasciatus</i>	BG	Good indication of extensive interspersion	Lagowski et al., 1973
<i>Drosophila melanogaster</i>	DG	Some sequence fraction may be interspersed, but probably repetitive sequences are longer than in sea urchin or toad	Kram et al., 1972 Wu et al., 1972 Manning, Schmid and N. Davidson, in prep.
Echuroid			
<i>Urechis caupo</i>	B*E*	Suggest interspersion	Authors' unpublished observations

Methods used in interspersion measurements:

- A Reassociation of long labeled fragments with short driver fragments at repetitive Cot.
- A* Extended to a wide range of measured fragment lengths.
- B Hyperchromicity of repetitive DNA fractions.
- B* Quantitated as a function of fragment length.
- C Rate of reassociation as a function of fragment length.
- D Electron microscopic measurements on reassociated repetitive DNA.
- E* Single-strand specific nuclease treatment of reassociated repetitive DNA fractions and measurement of size distributions of resistant regions.
- F Release of fragments containing only single-copy sequence by further shearing of fragments selected to contain repetitive sequence.
- G CsCl density of reassociated repetitive DNA.

another context below. In this work total sea urchin embryo mRNA in excess was reacted with labeled nonrepetitive DNA, and the hybridized DNA tracer was then reassociated with whole DNA (see Figure 5). The kinetics of the latter reaction were exactly those predicted for a single-copy DNA component, and little or no repetitive sequence reaction could be detected.

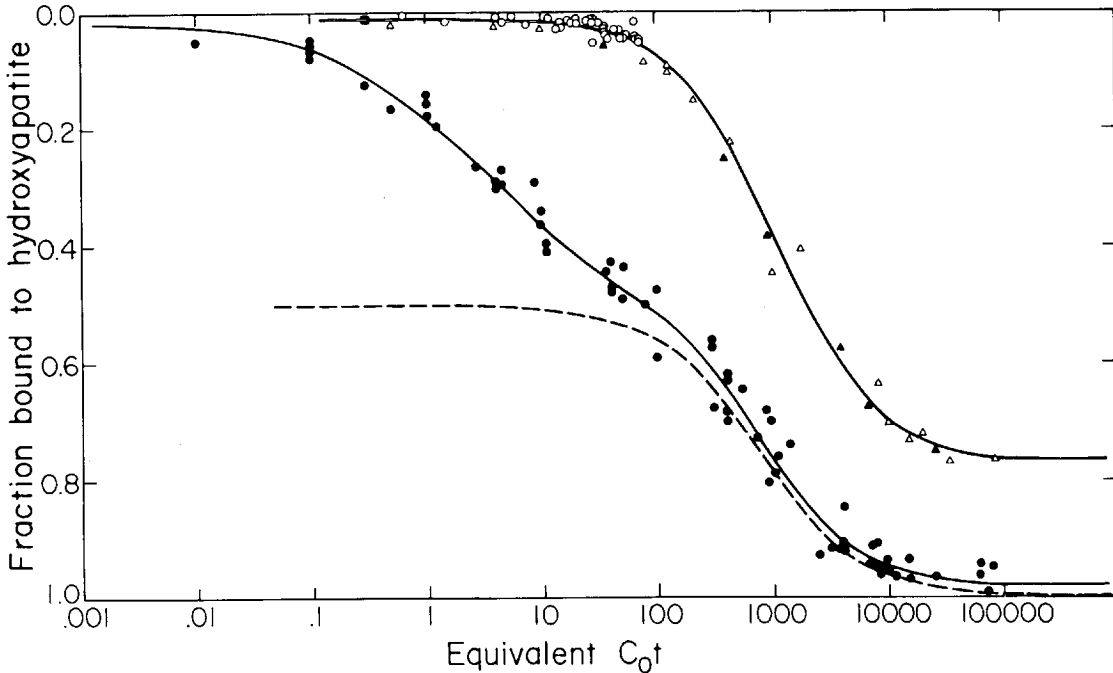


Figure 5. Reassociation of sea urchin DNA fractions (Galau et al., 1974). Reassociation of whole 450-nucleotide long DNA (●) at 60°C. Data are expressed in terms of equivalent C_0t s (i.e. C_0t corrected for the accelerating effects of $[Na^+]$ greater than 0.18 M). The curve is a least squares fit to the data utilizing three kinetic components (two of these are not shown). Reassociation of the nonrepetitive sequence component in the whole DNA is indicated by the dashed line. Its second order rate constant is 1.31×10^{-3} . Reassociation of nonrepetitive 3H -DNA (length about 220 nucleotides with a 10,000-fold excess of 450 nucleotide-long whole DNA (Δ) and with itself (\circ) are also shown. The C_0t s for the latter self reactions are corrected to the equivalent C_0t in whole DNA. The curve through the data is a least squares second order fit with a rate constant of 0.91×10^{-3} . Included are data describing reassociation of selected 3H -DNA extracted from RNA-DNA hybrids with whole 450 nucleotide long DNA (\blacktriangle) and with itself (\blacksquare). A fit to these data was virtually identical to that for the total nonrepetitive 3H -DNA-whole DNA reaction. The second order rate constant for the reaction of previously hybridized DNA with whole DNA was calculated to be 1.01×10^{-3} .

The sequence organization of mRNAs and of heterogeneous nuclear RNA (hnRNA) differ in at least one important respect. Though hnRNA consists mainly of single-copy sequence transcript, most of the molecules also contain covalently linked repetitive sequence elements. In contrast, mRNA does not. A summary of one kind of data supporting this assertion is presented in Table 3. Here two methods of hybridization are compared. The 'urea-phosphate hydroxyapatite' assay measures binding of hybridized mRNA in the absence of any RNase treatment, so that all parts of an RNA molecule containing a stable duplex region are bound. Thus the whole of an RNA fragment containing an interspersed repetitive sequence region would bind in urea-phosphate after a low C_{ot} incubation with DNA, while in an assay method including RNase treatment only that fraction of the molecule actually within duplex regions is scored as hybrid. A molecule which consists entirely of repetitive sequence should give the same answer in both assay systems. Table 3 shows that the fraction of HeLa mRNA which binds to hydroxyapatite at low C_{ot} (about 7%) is the same as that which is resistant to ribonuclease. Thus this fraction is transcribed entirely from repetitive sequences. The vast majority of the mRNA molecules appear to consist of non-repetitive sequence transcript, and since they do not bind to hydroxyapatite in urea-phosphate after low C_{ot} incubation with DNA, they must consist *entirely* of nonrepetitive sequence transcripts. In contrast, an average of one-third of each of the hnRNA fragments in Table 3 are transcribed from repetitive sequence and about two-thirds from nonrepetitive sequence. This follows from the fact that three times more hnRNA binds after low C_{ot} hybridization with ribonuclease treatment than is resistant to ribonuclease. The validity of these conclusions of course depends on lack of extensive RNA degradation during the hybridization procedure. This condition is met, as shown in the second and third columns of Table 3. The conclusion that hnRNA molecules contain interspersed repetitive sequences has also been reached for HeLa cell hnRNA by Darnell and his associates, as described elsewhere in this volume, and for ascites cell hnRNA by Holmes and Bonner (1974).

Size of structural gene sequences The question of structural gene length distribution can also be approached by analyzing the mRNA transcription products. For HeLa cell mRNAs, the best studied case, the weight average size of the RNA molecules is of the order of 2,700 nucleotides (see review in Davidson and Britten, 1973). However, the distribution of mRNA sizes is skewed and the median length of mRNA molecules, i.e. that length such that half the molecules are larger and half smaller, is around 1,400 nucleotides. This length includes about 200 nucleotides of poly(A). Similar mRNA size distributions are found in many animal cell types. Furthermore, it

TABLE 3
Hybridization properties of messenger and nuclear RNAs

RNA and DNA hybridized	Weight average size of RNA molecules ^a		C ₀ t	% of RNA hybridized ^b	
	Before hybridization	After hybridization		Urea-phosphate hydroxyapatite assay	RNase-sephadex assay
HeLa mRNA				5.6	7.6
with HeLa DNA ^c	2,000 nucleotides	2,000 nucleotides	40	7.0	7.1
Sea urchin hnRNA	3,000 nucleotides	1,100 nucleotides	40	22.4	8.4
with sea urchin DNA ^d				23.2	7.9
				24.6	8.5
				21.2	7.8

a RNA size determinations were made by formaldehyde sucrose gradient analysis.

b No increase in percent of RNA hybridized at low C₀t could be observed at increased DNA/RNA ratios. In the urea-phosphate hydroxyapatite assay samples are brought to 0.2 M phosphate buffer, 1% SDS, 8 M urea after incubation, and placed directly on hydroxyapatite columns in this buffer at 40°. Single-stranded RNA does not bind to such a column, while total RNA containing a hybridized portion does bind. The RNase-sephadex assay involves a high salt RNase treatment (0.24 M phosphate buffer) which degrades all RNA not duplexed with DNA. The samples are then placed over sephadex G200 columns which exclude hybrids and include RNA degradation products (see Smith et al., 1974, for further details).

c HeLa mRNA was prepared from a polysomal fraction. The DNA fragments measured 300 nucleotides in alkaline sucrose gradients. Data from Klein et al., 1974.

d *Strongylocentrotus purpuratus* hnRNA was prepared from gastrula nuclei. The DNA fragments measured 300 nucleotides in alkaline sucrose gradients. Data from Smith et al., 1974. Similar results could be expected with HeLa hnRNA, which has been shown to contain both repetitive and nonrepetitive sequence transcript, covalently linked (Darnell and Balint, 1970; Molloy et al., 1974).

seems clear that eukaryotic mRNAs are not in general polycistronic (Petersen and McLaughlin, 1973; reviewed by Davidson and Britten, 1973). A typical animal structural gene thus appears to exist in the DNA as a single nonrepetitive sequence. Subtracting the poly(A) sequence length, the median size of the distribution of structural gene sequence lengths would be about 1,200 nucleotides. The mean length (weighted average) would be about 2,500 nucleotides. Certain particular structural gene sequences are, of course, up to an order of magnitude larger than this, such as that coding for the huge silk fibroin protein (Suzuki and Brown, 1972). Others, such as the 500-600 nucleotide α -hemoglobin structural gene, are significantly smaller.

A significant though speculative conclusion is that the length distribution for interspersed single-copy sequence elements in sea urchin and *Xenopus* DNAs is very similar to the length distribution expected for structural gene sequences.

How many structural genes are there in the animal genome? To answer the question posed by this heading would require measurement of the number of different species of functional mRNA active throughout the life cycle. Even more onerous, such a measurement should in principle include all behavioral states for which the animal might require special information in the form of otherwise unused structural genes. No-one has yet approached such a formidable task. Until very recently, in fact, no measurement of the number of species of mRNA present at even a single stage of the life cycle has been available. The first observation of this kind was carried out in our laboratory by Galau et al. (1974) and is summarized in Figures 5 and 6.

Large amounts of mRNA were isolated from polysomes of sea urchin gastrulae. As noted above the method of extraction ensures that the only heterogeneous RNA sequences in the preparation are those of mRNA. The mRNA is then hybridized in excess with very small quantities of a highly labeled DNA fraction consisting mainly (>95%) of nonrepetitive sequences. The reassociation of the nonrepetitive tracer with whole sea urchin DNA (closed circles) is shown in Figure 5 (open symbols). It is clear that the nonrepetitive DNA fraction reassociates with the expected kinetics (see legend to Figure 5). Figure 6 shows the result of the reaction of mRNA with the nonrepetitive DNA. Since the mRNA is present in excess, its concentration does not change appreciably during the reaction, and the hybridization occurs with first order kinetics, at a rate determined by the concentration of hybridizing RNA sequences. Figure 6 illustrates the first order form of this hybridization reaction. At completion of the reaction about 1% of the total labeled DNA is hybridized, as indicated on the ordinate of Figure 6.

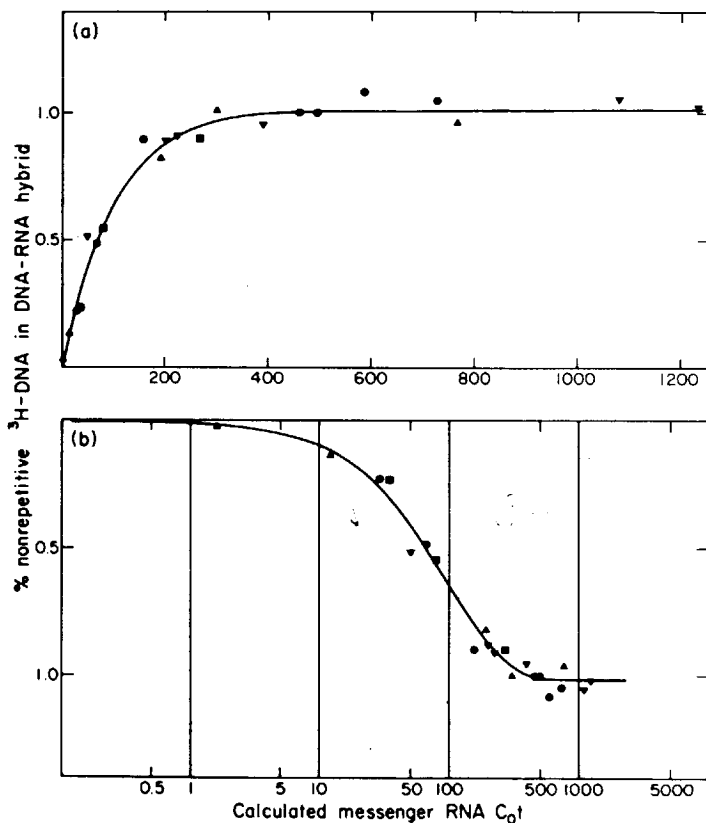


Figure 6. Hybridization of nonrepetitive ^3H -DNA with messenger RNA (Galau et al., 1974). Hybridization mixtures containing excess mRNA and the ^3H -labeled nonrepetitive DNA fraction whose reassociation with whole DNA is shown in Figure 5 were incubated to the RNA C_0t s shown. Four different preparations of mRNA were used, as denoted by the different symbols. Two methods of presenting the same data are shown. The curves were calculated according to a pseudo first order rate equation. Solid lines represent the least squares solution for the data. The first order rate constant of this fit is 0.010. The hybridization reaction extends from 0.0006 to 1.016% of the nonrepetitive ^3H -DNA. That is, the extent of the observed reaction is 1.01% of the ^3H -DNA in the experiment. RNA equivalent C_0t is denoted by RNA C_0t on the abscissa.

To determine whether hybridization had indeed occurred between mRNA and nonrepetitive DNA sequence, the hybridized DNA was extracted and its complementarity to the mRNA proved by a second reaction with mRNA. When the hybridized DNA was then reacted with whole DNA, it reassociated at exactly the rate characteristic of nonrepetitive sequence. These data are included in Figure 5 (solid triangles). Therefore, the amount of hybridized single-copy DNA at completion of the reaction of Figure 6 provides a direct measure of the mRNA

complexity. Taking into account that 75% of sea urchin DNA is present in nonrepetitive sequence (Graham et al., 1974) and that structural gene DNA is represented asymmetrically in polysomal mRNA, it is calculated that the 1% of the nonrepetitive DNA hybridized represents an mRNA complexity of about 17 million nucleotide pairs. If we assume a median length for structural gene sequences of about 1,200 nucleotides (see above), such an mRNA complexity is sufficient to comprise some 14,000 distinct structural genes. This value also gives an approximation of the number of diverse proteins being translated in the gastrula. Another conclusion derived from the data of Figure 6 is that probably 10%-30% of the total mRNA molecules include most of the diverse species of mRNA. That is, most of the mRNA population consists of a relatively small number of diverse species represented a relatively large number of times. This conclusion depends on the rate of hybridization, which is about 10 times slower than it would have been if all of the mRNA species were of the class whose hybridization is measured, assuming that the rate of hybridization (in excess RNA) approximates the rate of DNA-DNA duplex formation. If the RNA-DNA rate is slower, the fraction of the mRNA occupied by the diverse mRNA species class will be proportionately larger.

If as many as 14,000 structural gene products are required to program an embryo as simple as a 600 cell sea urchin gastrula, it would seem likely that far more exist in the genome as a whole. Thus, an appreciable fraction of the interspersed single-copy sequence elements in this genome could consist of structural gene sequences.

REPETITIVE SEQUENCE ELEMENTS IN THE IMMEDIATE VICINITY OF STRUCTURAL GENES

Here we ask whether single-copy structural gene sequences are indeed contiguous to the interspersed repetitive sequence elements in the genome. Our approach is to isolate that fraction of nonrepetitive DNA which is contiguous to repetitive sequence elements in the genome, and to compare the hybridization of mRNA with this DNA fraction to the hybridization obtained with whole DNA. In the following description we will refer to the preparation of DNA we studied as the 'repeat-contiguous' DNA fraction. These experiments will be reported in detail by Davidson, Hough, Klein and Britten (in preparation).

To obtain the repeat-contiguous DNA fraction, fragments of sea urchin DNA 2,000-3,000 nucleotides long were renatured to low C_{ot} in order to ensure duplex formation in repetitive sequence regions, but only repetitive sequence regions. An outline of the fractionations used is provided in Figure 7. As expected, the bound DNA fragments contain both repetitive and nonrepetitive sequence. Follow-

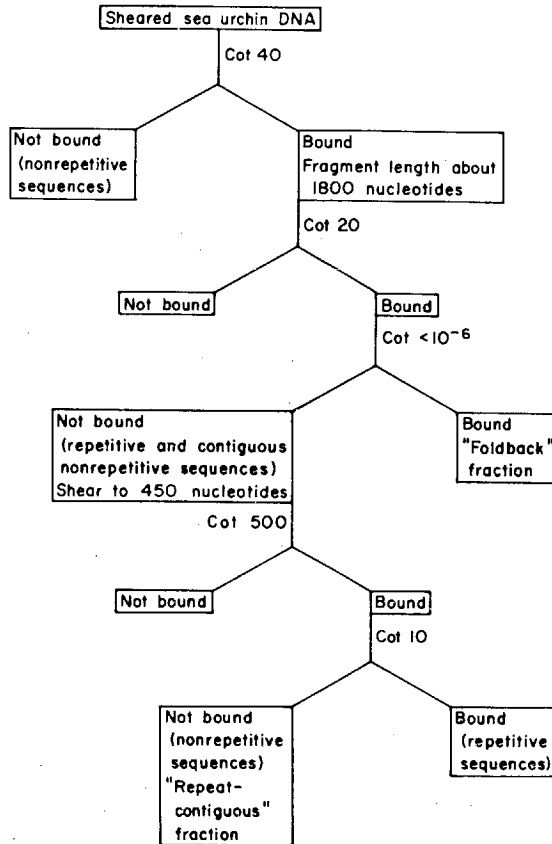


Figure 7. Preparation of a sea urchin DNA fraction containing nonrepetitive sequences which are contiguous to repetitive sequences in unfractionated DNA (Davidson, Hough, Klein and Britten, in preparation). The DNA measured 3,700 nucleotides in alkaline sucrose gradients before fractionation and about 1,800 nucleotides after the first hydroxyapatite fractionation and until the final shearing and fractionation step. Though only one fractionation is diagrammed, two successive low C_{OT} fractionations were actually carried out. After removal of fold-back fractions at C_{OT} 0.006 (i.e. intrastand duplex), the fragments were sheared to 450 nucleotides, reassociated to C_{OT} 10, and again fractionated in hydroxyapatite. DNA not bound after this C_{OT} 10 incubation consisted mainly of single-copy sequence. This fraction, termed the 'repeat-contiguous fraction'; was hybridized with labeled mRNA. Except for trace contaminants present at very low concentration repeat-contiguous DNA includes 20-30% of the total single copy complexity.

ing a preliminary fractionation at extremely low C_{OT} to remove the 'fold-back' DNA fraction (Figure 7); the fragments are sheared to 450 nucleotides, incubated to C_{OT} 10, and again fractionated on hydroxyapatite. The non-bound fraction is the 'repeat-contiguous fraction'. Though it contains trace amounts of most of the single-copy regions in the genome, the large majority of sequences present

in the repeat-contiguous DNA preparation represent only 20-30% of the total single-copy sequence. These are the sequences which were adjacent (i.e. within the range 0-1,800 nucleotides) to repetitive sequences in the starting DNA. The extent of hybridization of mRNA with this repeat-contiguous DNA fraction is compared with the hybridization of the same mRNA with whole DNA (sheared to the same fragment length). In whole DNA, of course, those single-copy sequences distant from repetitive sequence elements as well as those sequences contiguous to them are all proportionally represented.

Table 4 summarizes the results of this key experiment. The extent to which this mRNA preparation hybridized with whole DNA is about $65 \pm 5\%$ (i.e. terminal value). The extent of hybridization obtained with the repeat-contiguous DNA fraction was about $50 \pm 5\%$. If the mRNAs were derived from randomly distributed nonrepetitive sequences, only 12-20% hybridization would have been expected (i.e. $65\% \times 0.2-0.3$). Table 4 also presents calculations of the expected extent of hybridization based on models in which the repetitive sequence regions are located only at one end, or at both ends, of the structural gene sequences. It can be seen that the data are consistent with an organization in which 80-100% of the sequences coding for mRNA are located immediately next to repetitive sequence regions. It is clear that the data are *not* consistent with any model in which the location of structural genes is random with respect to the location of repetitive sequence elements, or in which structural genes are preferentially located in regions of the genome hundreds or thousands of nucleotides distant from repetitive sequences.

These important conclusions require that the mRNA hybridize with the major sequence classes in the repeat-contiguous DNA preparation rather than with trace contaminants representing more distant non-repetitive sequence regions. That this assumption is valid is shown a) by the fact that the kinetics of the hybridization reaction with repeat-contiguous DNA are just the same, relative to the driver DNA reassociation kinetics, as with whole DNA (where there are no 'rare' sequences), and b) by an experiment using the principle of Figure 4, in which the size of the particle containing the mRNA hybrids is determined by agarose gel filtration. Both forms of evidence demonstrate that the mRNA is hybridizing exclusively with the major sequence class in the repeat-contiguous DNA preparation rather than with rare contaminants. We conclude, therefore, at least in this animal genome, most or all structural gene sequences are immediately contiguous to interspersed repetitive sequence elements.

Though history may well prove us wrong, we believe that these observations are highly significant for the field of structural gene regulation. We have learned that the genome is highly ordered with

TABLE 4

Model calculations of sequence organization around structural genes from mRNA hybridization data ^a

I.	mRNA hybridization data. The values are subject to uncertainty in determining the exact termination of the hybridization reaction. We estimate this uncertainty to be about ± 0.05 . The mRNA fragments used had a modal length of about 550 nucleotides.	
	A. Fraction of mRNA hybridizing with whole DNA	.65
	B. Fraction of mRNA hybridizing with repeat-contiguous DNA	.50
II.	mRNA hybridization calculation. The fraction of mRNA which could hybridize with repeat-contiguous DNA is calculated from a typical mRNA size distribution curve, assuming that the DNA fraction includes 1,750 nucleotide pairs of non-repetitive sequence from each side of every interspersed repetitive sequence. ^b The calculated hybridization is normalized to a total possible hybridization of 0.65, as in I.A. above. In actuality, of course, the fragment length is not uniform and the modal fragment length is used here. Calculations based on this length give estimates of repeat-contiguous DNA complexity which agree with measurements.	
	A. If all structural genes are contiguous to repetitive sequences at one end	.50
	B. If all structural genes are contiguous to repetitive sequences at both ends	.62
	C. If structural genes are randomly distributed in the genome with respect to repetitive sequence elements. This calculation merely reflects the complexity of the repeat-contiguous DNA as measured in several ways.	.12-.22
III.	Estimations of the fraction of expressed gene sequences (mRNAs) which are contiguous to repetitive sequence elements by comparison of the hybridization data (I) with the model calculations (II). These estimates are subject to the uncertainties and assumptions noted above.	
	A. If structural genes have repetitive sequences at only one end. This represents a maximum estimate.	$\frac{(I_B).5}{(II_A).5} = 1.$
	B. If structural genes have repetitive sequences immediately adjacent to each end (a minimum estimate).	$\frac{(I_B).5}{(II_B).62} = 0.8$

^a Davidson, Hough, Klein, and Britten, in preparation.

^b The weight average mRNA size is 2,500 nucleotides; the largest mRNA molecules are over 10,000 nucleotides long. Given that the mRNA fragments require about 50 nucleotides in duplex to form stable hybrids, 500 nucleotides could extend beyond the length of the DNA. For purposes of calculation, the available length of DNA sequence available is assumed to include exactly 1,750 nucleotides of nonrepetitive sequence on each side of the repetitive sequence. The repeat-contiguous DNA was measured as an average of 1,800 nucleotides long in alkaline sucrose gradients, and 50 nucleotides of this is required for binding in the selection of the fraction.

respect to the spacing and interspersions of repetitive and nonrepetitive sequence elements. The last experiments place the structural gene in this pattern of sequence arrangement. We note that this evidence fulfills a crucial prediction of the Britten-Davidson model of gene regulation (Britten and Davidson, 1969; Davidson and Britten, 1973). In this model the co-ordinated transcription of sets of structural genes is organized through the interaction of sequence-specific hybrids with homologous repetitive sequence elements. These 'receptor sequences' are predicted to be located contiguous to the 5'-end of the structural genes (with respect to the mRNA). It remains to be seen whether the repetitive sequence elements in the genome are present in this location and more importantly, whether they indeed play a role in the activation of transcription.

ACKNOWLEDGMENT

This work was supported by USPHS grant No. HD 05753.

REFERENCES

- Britten, R.J. 1972. DNA sequence interspersions and a speculation about evolution. *In* Evolution of Genetic Systems (H.H. Smith, ed.) p 80. Gordon and Breach, New York.
- Britten, R.J. and E.H. Davidson. 1969. Gene regulation for higher cells: a theory. *Science* 165: 349.
- Britten, R.J. and J. Smith. 1970. A bovine genome. *Carnegie Inst. Wash. Year Book* 68: 378.
- Darnell, J.E. and R. Balint. 1970. The distribution of rapidly hybridizing RNA sequences in heterogeneous nuclear RNA and mRNA from HeLa cells. *J. Cell. Physiol.* 76: 349.
- Davidson, E.H. and R.J. Britten. 1973. Organization, transcription and regulation in the animal genome. *Quart. Rev. Biol.* 48: 565.
- Davidson, E.H., B.R. Hough, C.S. Amenson and R.J. Britten. 1973. General interspersions of repetitive with nonrepetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* 77: 1.
- Davidson, E.H., B.R. Hough, M.E. Chamberlin and R.J. Britten. 1971. Sequence repetition in the DNA of *Nassaria (Ilyanassa) obsoleta*. *Develop. Biol.* 25: 445.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *In* Methods in Enzymology (L. Grossman and K. Moldave, ed.) Vol. 21, p.413. Academic Press, New York.
- Galau, G.A., R.J. Britten and E.H. Davidson. 1974. A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. *Cell* 2: 9.
- Goldberg, R.B., G.A. Galau, R.J. Britten and E.H. Davidson. 1973. Non-repetitive DNA sequence representation in sea urchin embryo

- messenger RNA. *Proc. Nat. Acad. Sci. USA* 70: 3516.
- Graham, D.E., B.R. Neufeld, E.H. Davidson and R.J. Britten. 1974. Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. *Cell* 1: 127.
- Holmes, D.S. and J. Bonner. 1974. Interspersion of repetitive and single-copy sequences in nuclear ribonucleic acid of high molecular weight. *Proc. Nat. Acad. Sci. USA* 71: 1108.
- Klein, W.H., W. Murphy, G. Attardi, R.J. Britten and E.H. Davidson. 1974. Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. *Proc. Nat. Acad. Sci. USA* 71: 1785.
- Kram, R., M. Botchan and J.E. Hearst. 1972. Arrangement of the highly reiterated DNA sequences in the centric heterochromatin of *Drosophila melanogaster*. Evidence for interspersed spacer DNA. *J. Mol. Biol.* 64: 103.
- Lagowski, J.M., M.-Y.W. Yu, H.S. Forrest and C.H. Laird. 1973. Dispersity of repeat DNA sequences in *Oncopeltus fasciatus*, an organism with diffuse centromeres. *Chromosoma* 43: 349.
- Molloy, G.R., W. Jelinek, M. Salditt and J.E. Darnell. 1974. Arrangement of specific oligonucleotides within poly (A) terminated HnRNA molecules. *Cell* 1: 43.
- Peterson, N.S. and C.S. McLaughlin. 1973. Monocistronic messenger RNA in yeast. *J. Mol. Biol.* 81: 33.
- Rice, N. 1971. Thermal stability of reassociated repeated DNA from rodents. *Carnegie Inst. Wash. Year Book* 69: 472.
- Smith, M.J., B.R. Hough, M.E. Chamberlin and E.H. Davidson. 1974. Repetitive and nonrepetitive sequence in sea urchin hnRNA. *J. Mol. Biol.* 85: 103.
- Suzuki, Y. and D.D. Brown. 1972. Isolation and identification of the messenger RNA for silk fibroin from *Bombyx mori*. *J. Mol. Biol.* 63: 409.
- Wilson, D. and C.A. Thomas. 1973. Hydroxyapatite chromatography of short double-helical DNA. *Biochim. Biophys. Acta* 331: 333.
- Wu, J.-R., J. Hurn and J. Bonner. 1972. Size and distribution of the repetitive segments of the *Drosophila* genome. *J. Mol. Biol.* 64: 211.

An analysis of *Drosophila* DNA by restriction endonucleases

C. A. Thomas Jr, D. H. Hamer, R. E.
Pyeritz, G. P. Holmquist, K. Shastri
and S. Manteuil

Drosophila melanogaster DNA, together with λ and *Escherichia coli* DNAs as controls, was digested with a series of three different restriction endonucleases and the resulting distributions of segments were characterized. More than 85% of the *D. melanogaster* DNA was found in a broad distribution of segment lengths consistent with random location of recognition sites. However, some DNA was spared and recovered in very long ($\geq 20,500$ bp) segments. These segments proved to be mostly simple sequence DNA. No complex, long spared segments could be found in Hind and Hae digests, while 50% of the spared EcoRI segments had a complexity exceeding that of *E. coli* DNA. These data do not support the hypothesis that chromomeres contain long regions of purely tandemly repeating sequences.

About two years ago we completed a group of experiments based on the formation of 'folded rings' from DNA fragments obtained from *Drosophila* and other species (see Thomas et al., 1973). The results could be accounted for by a very simple model of the chromosome. This model states that the chromatid is a simple Watson-Crick double helix, the nucleotide sequences of which are organized into thousands of specific regions, called 'g-regions'. These g-regions are thought to be 3,000-150,000 base pairs (bp) (1-50 μ) in length and each composed of a distinctive high-fidelity, tandemly repeating sequence. The repeat length, called s in our analysis, is thought to be 600-6,000 bp (0.2-2 μ) with a number average near 1,000 bp. Our experiments indicated that about one half of the DNA resided within g-regions, and the other half in some other kind of sequences which were not detected by our experiments. A diagram of this model is shown in Figure 1. The evidence supporting this model has been published and will not be reviewed here.

This model makes some specific predictions that we have been attempting to test during the past year. The experiments described here have to do with the molecular weight distributions of DNA segments that are produced when high molecular weight *Drosophila* DNA is digested with three different restriction endonucleases.

Although it is unfashionable to do so, it must be said at the outset that the data offer no support for the g-region model as described.

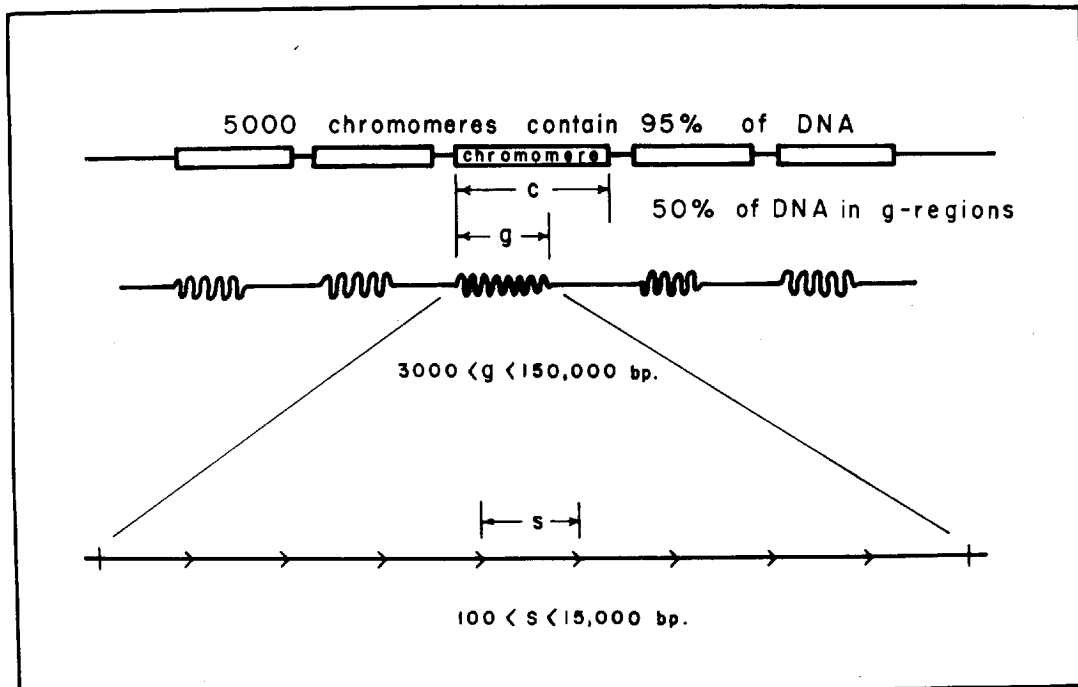


Figure 1. The g-region model for the *Drosophila* chromosome. Approximately one half of the DNA in each chromomere is pictured to be in a region g bp in length composed of a high fidelity tandemly repeating sequence of length s .

Restriction endonucleases have been discovered in a wide variety of microorganisms. They have the remarkable ability to recognize a special sequence of 4-8 nucleotides and to introduce two phosphodiester chain breakages (one in each chain) in or near this specific recognition sequence. These nucleases introduce a limited number of double-chain breakages into a viral DNA molecule to produce a collection of equimolar segments of defined length. These segments can be separated from each other on the basis of length by electrophoresis through agarose gels and detected by their radioactivity or by fluorochroming with ethidium bromide.

In this work we have employed three restriction endonucleases: $EcoR_I$, obtained from *E. coli* carrying a drug resistance plasmid (Yoshimori, 1971); $Hind$, obtained from *Haemophilus influenzae* strain Rd, and consisting of a mixture of two different restriction endonucleases (Smith and Wilcox, 1970; Danna et al., 1973); and Hae , one of the restriction endonucleases found in *Haemophilus aegyptius*

(Middleton et al., 1972). The recognition sequences of these three nucleases have been determined on bacteriophage DNA (Table 1), and we assume that the same recognition sequences are cut in eukaryotic DNA. On the basis of the known recognition sequence and the average nucleotide composition of the DNA, it is possible to calculate the expected density of recognition sites, p , in a given DNA. As will be illustrated shortly, p can also be determined experimentally. Calculated and found values of $1/p$, the number average chain length of the segments, are given in Table 1. As expected, the longer the recognition sequence, the less frequently it is found, and the longer the average chain length of the collection of segments.

Some g-regions should be totally unbroken by a given restriction endonuclease If half the DNA were in g-regions, which themselves were composed of perfect tandem repeats of length s , then by chance some of these sequences s bp long should *not* contain the recognition sequence of the endonuclease. In this case, the entire g-region should be spared, and some of these regions should correspond to quite long segments of more than 15,000 bp (5μ). If the recognition sites were distributed strictly randomly, then the fraction of DNA in spared g-regions, F_0 , should be

$$F_0 = \gamma e^{-ps} \quad (1)$$

where γ is the fraction of all the DNA in such regions estimated to be about 50%. For example, if $s = 1,000$, and the values of $1/p$ are those shown in Table 1, then the expected fraction of DNA in spared g-regions would be 41% for an $EcoR_I$ digest, 18% for a $Hind$ digest and 2.9% for a Hae digest.

On the other hand, if this g-region model had no validity at all, and recognition sites were distributed randomly throughout the DNA, then only a very small percentage of long segments would be spared. In this model the fraction of DNA that is considered spared will depend on the length chosen to define such segments. We know from the Kuhn approximation of the equations of Montroll and Simha (1940) that the weight fraction, w_t , of segments exactly t bp long is

$$w_t = p^2 t(1-p)^{t-1} \quad (2)$$

Integrating (2) from $t = 0$ to t , the expression becomes

$$\int_0^t w_t dt = 1 - (1+pt)e^{-pt} \quad (3)$$

Hence the fraction of DNA remaining longer than t bp is

$$\int_t^\infty w_t dt = (1+pt)e^{-pt} \quad (4)$$

TABLE 1
The density of recognition sequences

Endo-nuclease	Restriction site(s)	1/p (base pairs)			
		Calculated <i>E. coli</i> (50% GC)	Calculated <i>D. mel.</i> (42% GC)	<i>E. coli</i>	Found <i>D. mel.</i>
EcoR _I	5' (A/T)G+A A T T C(A/T)	4,096	3,206	5,000±1,000 (100%)	2,500±500 (90%)
	(A/T)C T T A A↑G(A/T)				
Hind	5' G T Y+U A C	1,024	1,078	1,000±200 (100%)	1,000±200 (88%)
	C A U+Y T G	819	807		
	5' A↑A G C T T				
	T T C G A↑A	4,096	3,206		
Hae	5' G G+C C	256	514	350±70 (100%)	350±70 (86%)
	C C↑G G				

The three different restriction endonucleases are given by their abbreviations. This is followed by the recognition sequence and cleavage points. The calculated reciprocal of the density of recognition sites (1/p) or the number average chain length is followed by the value observed from Figure 4. In brackets the percentage of the DNA having an apparently random distribution of recognition sites is given.

For the molecular weight limit chosen ($t = 20,500$ bp) this expression predicts a fraction of surviving segments that is less than $1/100$ the value expected from the g -region model, given by equation (1). Thus, the two models make sharply contrasting predictions and are susceptible to an experimental test.

Size distributions of λ , *E. coli* and *D. melanogaster* restriction segments To approach this problem we have electrophoresed terminal digests of *D. melanogaster* tissue culture cell DNA, separately and together with co-digested λ and *E. coli* DNA, through agarose gels. Figure 2 shows terminal digests of λ , *E. coli* and *D. melanogaster* DNA made with $EcoR_I$, Hind and Hae respectively. In this series the concentration of agarose (1 or 2.4%) was sufficiently high to ensure that all the segments would be retained on the gel. In Figure 2a one may see five of the six λ segments, the suggestion that *E. coli* DNA is broken into a large number of segments that are partly resolved from the general blur, and finally the *D. melanogaster* distribution which is not resolved into individual zones with the exception of some very high molecular weight material ($>20,000$ bp) which migrates at the limiting rate. Figure 2b shows a comparable series for Hind segments. Here λ DNA is broken into a larger number of segments; *E. coli* DNA produces a collection of closely spaced zones, and *D. melanogaster* DNA produces an unresolved distribution, again with some very long segments moving at the limiting rate, which on this gel is 4,500 bp. In Figure 2c one sees comparable behavior for the Hae segments, except that they are smaller and more numerous. Again a significant portion of the *D. melanogaster* DNA remains longer than 4,500 bp.

In order to quantitate these observations, all three DNAs were digested together in the same tube: a few μg of unlabeled λ , ^{32}P -*E. coli*, and ^3H -*D. melanogaster*. The digestion was carried out with sufficient enzyme and for a sufficient length of time to generate the characteristic λ segment pattern which was fluorochromed and photographed. These gels were sliced and the two radiolabels counted. Knowing the lengths of the various λ segments, a calibration curve was constructed of log segment length versus slice number. The results shown in Figure 3 display a general distribution of segment lengths for both *E. coli* and *D. melanogaster* DNA with the notable difference that digests of *D. melanogaster* DNA contain some long segments migrating at the limiting rate. Therefore, two problems present themselves; what can be learned from these distributions of short segments; and what is the very high molecular weight material that migrates in an unresolved manner near the top of the gel?

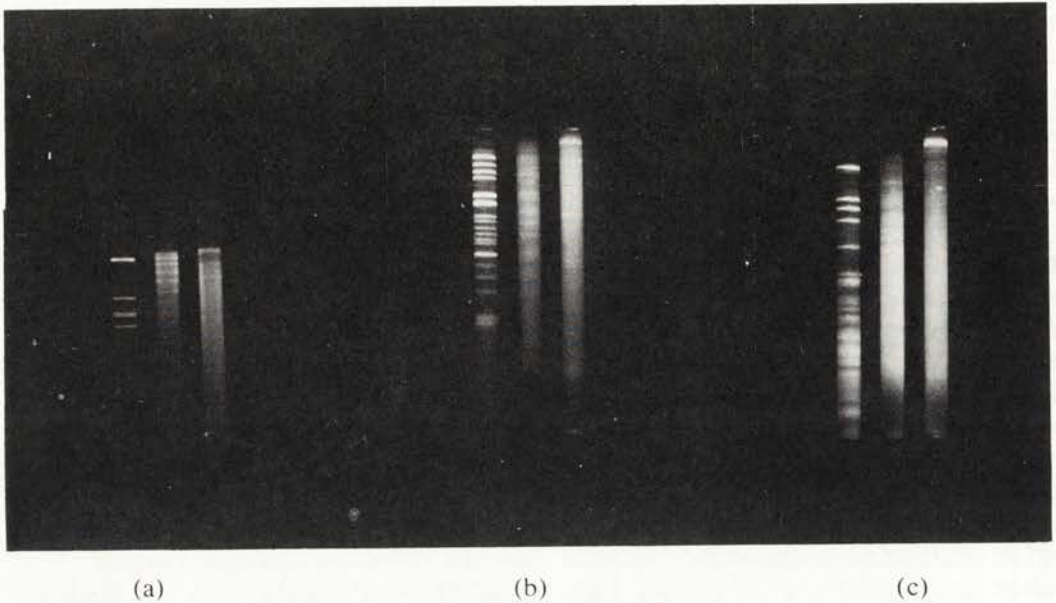


Figure 2. Electrophoresis of λ , *E. coli* and *D. melanogaster* restriction segments through concentrated agarose gels. Total cell DNA was isolated from *D. melanogaster* Schneider line II tissue culture cells (Schneider, 1972) and from *E. coli* BB thy⁻ by lysis with sarkosyl and pronase followed by equilibrium centrifugation in Cs_2SO_4 gradients. Bacteriophage λ was prepared by induction of the temperature sensitive lysogen M5107 ($\lambda\text{c1857S7}$) and extracted with phenol and chloroform. Restriction endonucleases EcoRI (Yoshimori, 1971) and Hae (endonuclease Z of Middleton et al., 1972) were purified by published methods with some modifications (Hamer and Thomas, in preparation). Restriction reactions contained 40 to 100 $\mu\text{g/ml}$ DNA, 6 to 10 mM MgCl_2 , and a 2.5-fold excess of endonuclease. After two hours at 37° , an additional 2.5-fold excess of endonuclease was added, and incubation was continued for 2 hours at 37° . Addition of more endonuclease and further incubation did not alter the gel patterns. Reactions were stopped by adding 0.1 volume of 2% sodium dodecyl sulfate, 0.05% bromophenol blue, 0.1 M EDTA, pH 8.7. Samples of 20 to 50 μl were electrophoresed through 85 x 6 mm agarose gels for 4 to 6 hours at room temperature at a constant current of 2 mAmps per gel. The buffer was 40 mM Tris, 5 mM sodium acetate, 1 mM EDTA, pH 7.75. Gels were stained with 0.5 $\mu\text{g/ml}$ ethidium bromide, illuminated with an ultraviolet lamp (long wavelength model, UV Products) and photographed through a yellow barrier filter and a Kodak #22 wratten filter. (a) EcoRI digests of (from left to right) λ (0.45 μg), *E. coli* (1.2 μg) and *D. melanogaster* (1.2 μg) DNA. 1% agarose gels. (b) HindIII digests of λ (1.2 μg), *E. coli* (2.4 μg) and *D. melanogaster* (2.4 μg) DNA. 2.4% agarose gels. (c) HaeIII digests of λ (1.2 μg), *E. coli* (2.4 μg) and *D. melanogaster* (2.4 μg) DNA.

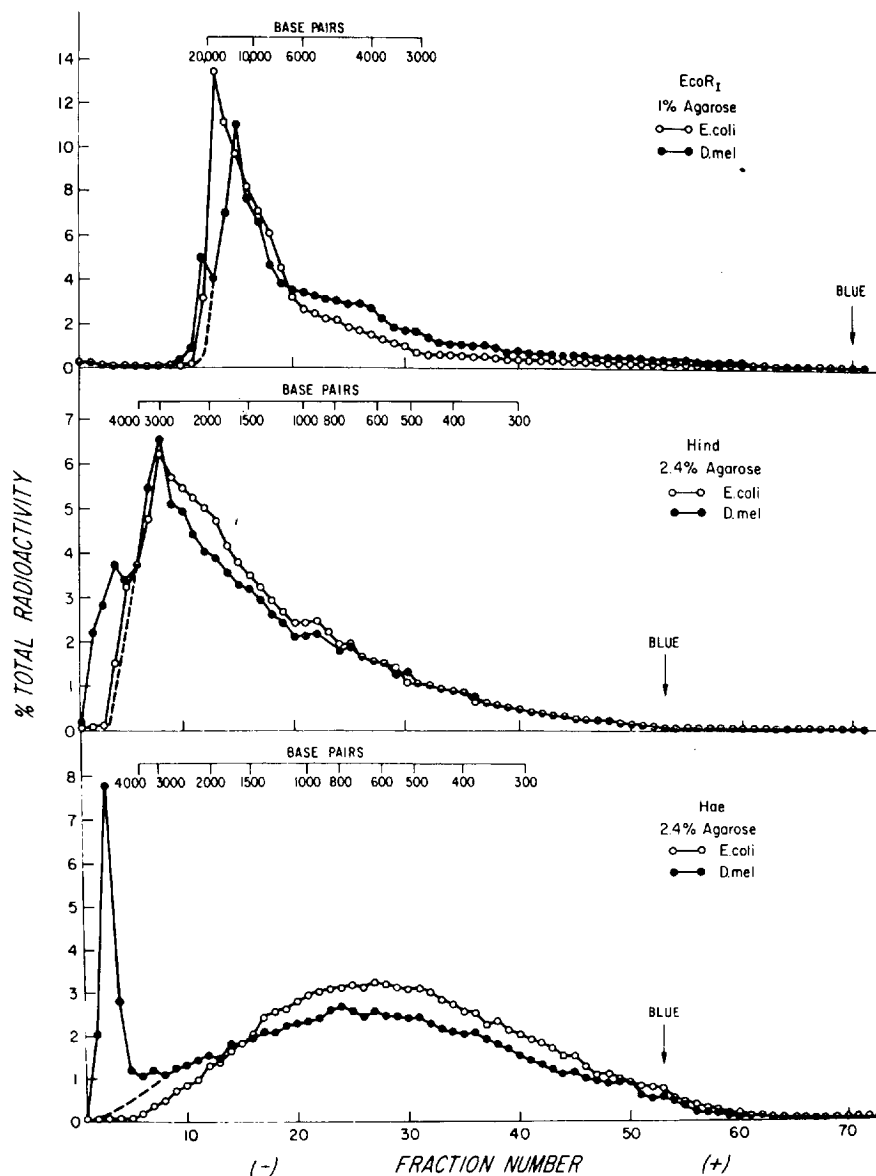


Figure 3. Quantitative determination of *E. coli* and *D. melanogaster* segment size distributions. *D. melanogaster* ^3H -DNA and *E. coli* ^{32}P -DNA were isolated as described in the caption to Figure 2 and examined by sedimentation through 5 to 25% neutral sucrose gradients (SW50.1 rotor, 49,000 rpm, 90 min). Both preparations sedimented 1.5 to 2 times faster than admixed phage T_7 DNA. Restriction reactions contained 1.2 μg of λ DNA, 0.035 μg of *E. coli* ^{32}P -DNA (1.3×10^6 cpm/ μg), 0.35 μg of *D. melanogaster* ^3H -DNA (1.9×10^5 cpm/ μg), and excess endonuclease. After electrophoresis, gels were stained with ethidium bromide and photographed to determine the migration rates of the λ segments, which have been sized by co-electrophoresis with T_7 /Hae and SV40/Hind segments (Hamer and Thomas, in preparation). In each reaction the λ DNA was terminally digested. Gels were cut into 1 mm slices which were solubilized in Soluene fluor and counted for ^{32}P and ^3H . The recovery of radioactivity from these gels was the same as from a gel containing undigested DNA, indicating that loss of small segments was negligible.

To approach the first problem, one may ask if the distribution of segment lengths that is observed is consistent with the *random* location of recognition sites. To test this, the data from Figure are displayed in an integral form plotted against $\log t$ (Figure 4). The smooth lines are plots of equation (3) using those values of p , the density of restriction sites, giving the best fits. In order

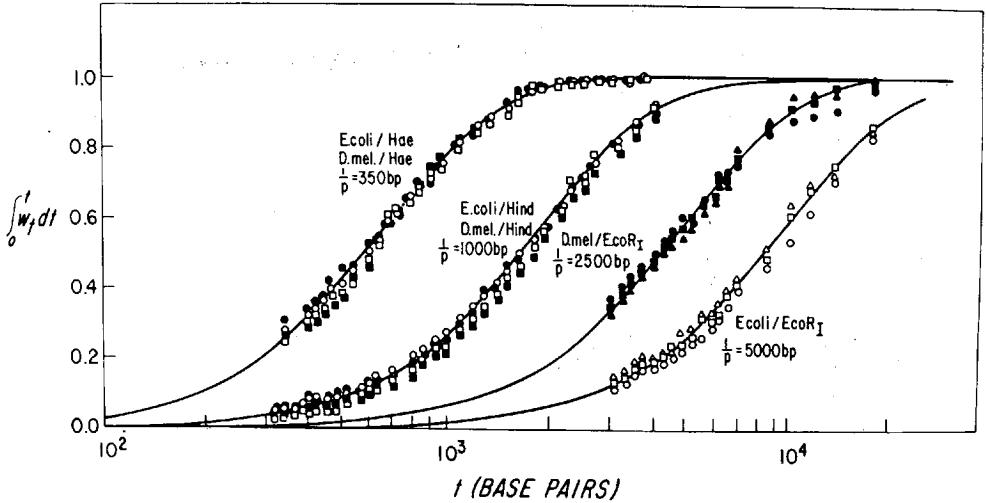


Figure 4. Cumulative size distributions of *E. coli* and *D. melanogaster* restriction segments. Calibration curves of $\log \lambda$ segment length versus slice number were constructed for each distribution shown in Figure 3. These curves were used to assign molecular weight limits to each gel slice falling between the largest and smallest visible λ segments. The weight fraction of restricted DNA having length less than or equal to t , $\int_0^t w_t dt$, was determined by summing the fraction of the total radioactivity present in all the gel slices from the bottom of the gel to the slice containing segments of maximum length t . In this analysis, we neglected the high molecular weight *D. melanogaster* segments by replacing the data points with a smooth curve (dotted line in Figure 3) for the fractions at the top of the gel. The smooth lines are plots of equation (2) using those values of p giving the best fit to the data. \circ, \square, Δ , *E. coli* segments; $\bullet, \blacksquare, \blacktriangle$, *D. melanogaster* segments.

to display the *D. melanogaster* DNA on this graph, the long segments were subtracted by drawing the dotted lines shown in Figure 3 parallel to the *E. coli* distributions. After making this correction *E. coli* and *D. melanogaster* DNAs have the same density of restriction sites for Hind ($1/p = 1,000$ bp) and for Hae ($1/p = 350$ bp). However, EcoRI sites are less frequent in *E. coli* DNA ($1/p = 5,000$ bp) than in *D. melanogaster* DNA ($1/p = 2,500$). This agreement with equation (3) does not mean that the recognition sites are located at random (God forbid!) but merely that the segment distribution

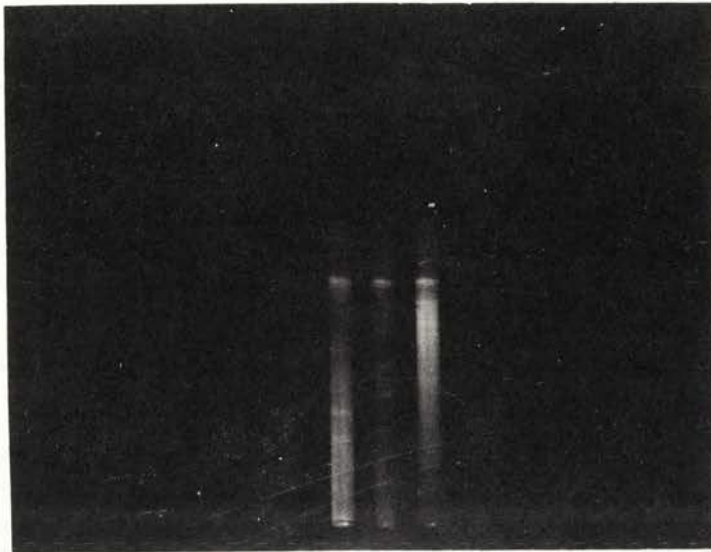
that results is indistinguishable from a random distribution which can be described by a single value of $1/p$. These values are listed in Table 1. From them we can state that the abundance of the very long *D. melanogaster* segments is more than 100 times that expected for a random distribution of recognition sites, as seen in *E. coli* DNA. (As an exercise to illustrate this point, the reader is encouraged to break this written text at any arbitrary letter - space, comma or other symbol - assuming it to be a 'recognition site'. Then plot the resulting text segments in the manner suggested by Figure 4 and assign a value of $1/p$ to the chosen letter!)

In order to study these long segments more quantitatively we have focused attention on segments equal to or longer than 20,500 bp, the length of the longest EcoRI segment of λ DNA, which serves as a convenient marker. Segments of this length can be clearly separated from the smaller segments by electrophoresis through dilute (0.85%) agarose gels. An example of such gels is shown in Figure 5. Such gels were sliced and the fraction of radioactivity migrating a distance equal to or less than the distance travelled by the 20,500 bp marker was determined. The results are shown in Table 2 where it can be seen that EcoRI spares 10%, Hind 5% and Hae 5% of the *D. melanogaster* DNA in the form of long segments. In the same gels no long segments can be found in Hind and Hae digests of *E. coli* DNA, but EcoRI spares 9.4% in long segments. Clearly, this is in accord with the expectation based on Figure 1 and equation (1). Therefore, it was of some interest to characterize these spared segments more carefully.

Sensitivity of spared segments to digestion by multiple endonucleases

A first step in the characterization of these spared segments can be found in Table 2. If the spared segments were a mixture of different tandemly repeating sequences of significant repeat length (say $s = 1,000$ bp), then one might expect that a smaller fraction would be spared by two endonucleases, and a still smaller fraction by three. If 5% were spared by one endonuclease, perhaps 5% of 5% or 0.25% would be spared by two, etc. However, this does not appear to be the case: about 5% is spared by any two or three endonucleases. The results shown in Table 2 demonstrate that exactly the same sequences are spared by Hind and by Hae, and that these sequences account for one half of the material spared by EcoRI. This means that the spared DNA is probably not a very complex mixture of sequences, or if so, the sequences themselves cannot be very complex.

Isopycnic sedimentation of spared segments Figure 6 shows how these spared segments behave in CsCl and actinomycin D-CsCl banding experiments. As can be seen the spared segments distribute themselves in



(a) (b) (c)

Figure 5. Electrophoresis of *D. melanogaster* restriction segments through dilute agarose gels. High molecular weight *D. melanogaster* DNA (1.5 μ g per reaction) was terminally digested with each endonuclease and electrophoresed through 0.85% agarose gels. On these gels, the largest λ EcoRI segment (20,500 bp) migrated 10% further than the limiting rate, while segments smaller than 800 bp ran off the bottom of the gel. Spared segments appear as a fluorescent band at the limiting rate. (a) Hind digest. (b) Hae digest. (c) EcoRI digest.

a number of density classes, four in the case of actinomycin D-CsCl gradients. These classes correspond exactly to the highly repeated, simple sequence DNAs previously described by Peacock et al. (1974) and named 1.705, 1.688, 1.686 and 1.672 g/cc satellites. In the case of EcoRI, about one half the radiolabel co-distributes with the main band. In contrast, something less than 10% of the Hind and Hae spared segments is found at the density of the main band. It appears that the spared segments are mostly DNAs of very simple sequence.

Association rates of spared segments This general picture is confirmed by association rate experiments. The spared segments were mixed with *E. coli* DNA of alternate radiolabel, fragmented to about 500 bp by sonication, denatured and allowed to associate for various periods, then passed over hydroxyapatite. The results in Figure 7 show that more than 90% of the Hae spared DNA associates 40,000 times

TABLE 2

Percent of terminal digest $\geq 20,500$ bp

Endonuclease	<i>E. coli</i> (^{32}P)	<i>D. mel</i> (^3H)
EcoR _I	9.4	10.7
Hind	<0.1	5.4
Hae	<0.1	5.0
EcoR _I plus Hind	-	5.1
EcoR _I plus Hae	-	5.3
Hind plus Hae	-	5.2
EcoR _I plus Hind plus Hae	-	5.0

λ , *E. coli*, and *D. melanogaster* DNA were codigested terminally with the endonucleases shown and electrophoresed through 0.85% agarose gels. The fraction of isotopes migrating at a rate equal to, or slower than, the 20,500 bp λ segment is shown. Approximately the same percentages of spared long segments are obtained by ethidium bromide fluorescence. Due to their low (GC) content the ^3H percentages could be at most 1.3 x higher than the percentage of DNA of average composition.

faster than *E. coli* DNA with a $C_{0t_{1/2}}$ of about 10^{-4} Ms/l. The EcoR_I spared DNA seems to be composed of two species: one half that associates at the same rapid rate as the Hae spared segments, and the other half that associates very slowly and probably corresponds to the spared segments that codistribute with the main band DNA in the sedimentation equilibrium experiments.

In situ hybridization to polytene chromosomes In order to complete the picture, the chromosomal location of the spared segments was determined by *in situ* hybridization. EcoR_I, Hind and Hae spared segments were used as templates for *in vitro* synthesis of high specific activity ^3H -RNA, which was purified and annealed to salivary gland squashes by established procedures (Gall and Pardue, 1971). Under our conditions, RNA made from *E. coli* DNA showed no specific association with the chromosomes. As expected, RNA made from Hind and Hae spared *D. melanogaster* DNA segments annealed almost exclusively with the chromocenter (Figure 8a, b). In contrast, RNA made from EcoR_I spared segments becomes associated with both the chromocenter and with all of the arms (Figure 8c). This finding is to be expected in view of the fact that one half of this DNA codistributes with the main band, and one half is slowly associating.

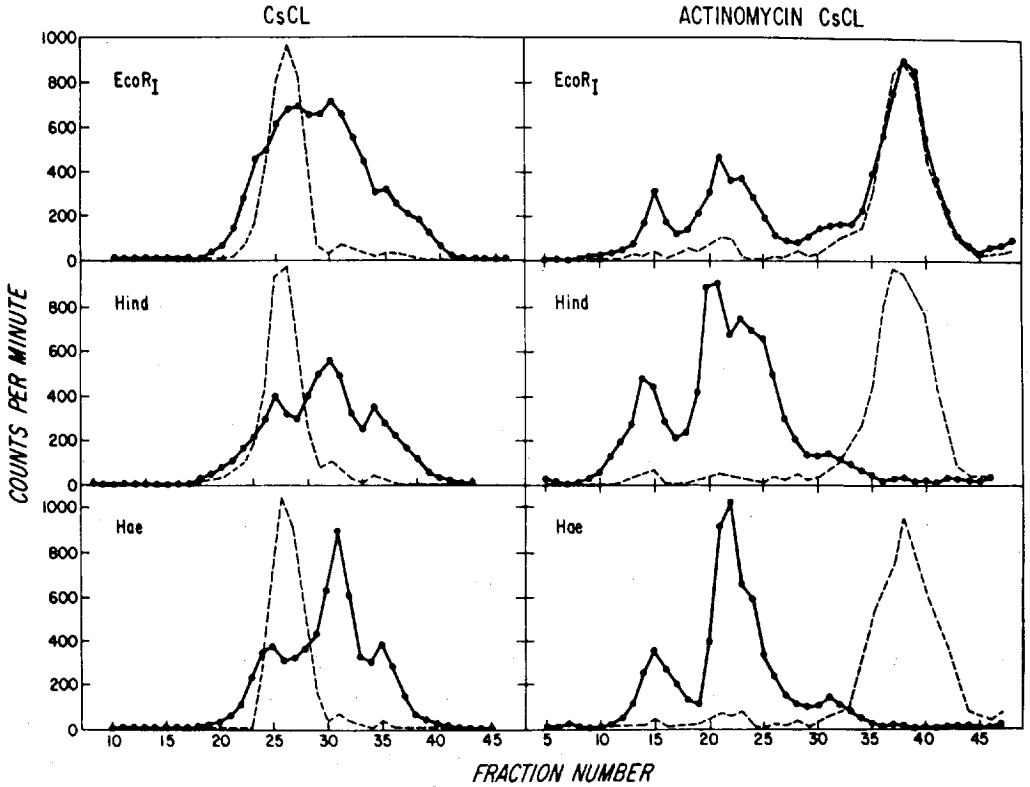


Figure 6. CsCl and actinomycin D-CsCl gradients of spared segments. *D. melanogaster* ³H-DNA was terminally digested with each restriction endonuclease and electrophoresed through preparative 0.85% agarose gels. After staining with ethidium bromide, ≥20,500 bp segments were eluted, extracted with isoamyl alcohol, and purified by hydroxyapatite chromatography. CsCl gradients (8 ml) had an initial density of 1.70 g/cc and were centrifuged at 35,000 rpm for 76 hours in a Beckman 40 rotor at 18°C. Actinomycin D-CsCl gradients (10 ml) contained 40 μg of actinomycin D and CsCl to give an initial density of 1.66 g/cc and were centrifuged at 45,000 rpm for 50 hours in a Beckman Ti60 rotor at 18°C. Gradients were collected from the bottom, and densities were estimated from refractive index measurements on every fifth fraction. ●—● spared segments (³H); - - - sheared, total *D. melanogaster* DNA (³²P).

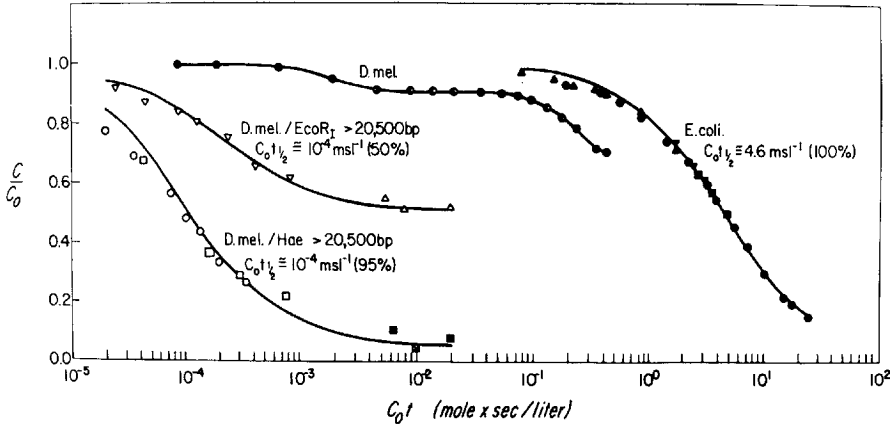


Figure 7. Association rates of spared segments. ^3H -labeled EcoR_I and Hae -spared segments were isolated as described in the legend to Figure 6 and sonicated to an average single chain length of approximately 500 nucleotides. The sonicated segments had the same average s values in alkaline sucrose gradients as sonicated *E. coli* ^{32}P -DNA, which was employed as an internal standard in some of the reassociation reactions. DNA in 0.14 M sodium phosphate buffer (NaPB), 0.1 mM EDTA, pH 6.8, was denatured at 100°C for 4 min, then immediately shifted to a $50.0 \pm 0.1^\circ\text{C}$ water bath. At appropriate intervals, samples were diluted with 0.14 M NaPB, loaded onto a water-jacketed column containing 1 ml packed volume of hydroxyapatite at 50°C , washed with 10 ml of 0.14 M NaPB, and eluted with 10 ml of 0.40 M NaPB. Separation of double-stranded from single-stranded DNA was quantitative. The fraction, B_0 , of the DNA binding due to intramolecular reassociation was determined by denaturing and loading the DNA at low ionic strength (0.002 M NaPB). The fraction, C/C_0 , of DNA remaining denatured at time t (in seconds) was calculated as $C/C_0 = (1 - B_t)/(1 - B_0)$, where B_t is the fraction of DNA bound in 0.14 M NaPB at time t . ●, *E. coli* (20.2×10^{-5} M, $B_0 = 0.000$) plus ○, *D. melanogaster* (4.81×10^{-6} M, $B_0 = 0.023$); ■, *E. coli* (11.9×10^{-5} M) plus □, *D. melanogaster/Hae* $\geq 20,500$ bp segments (3.03×10^{-7} M, $B_0 = 0.063$); △, *E. coli* (11.9×10^{-5} M) plus △, *D. melanogaster/EcoR_I* $\geq 20,500$ bp segments (3.92×10^{-7} M, $B_0 = 0.086$); ▲, *E. coli* (11.8×10^{-5} M); ⊙, *D. melanogaster* (1.00×10^{-6} M); ⊚, *D. melanogaster* (1.03×10^{-7} M); □, *D. melanogaster/Hae* $\geq 20,500$ bp segments (6.48×10^{-8} M); ○, *D. melanogaster/Hae* $\geq 20,500$ bp segments (5.50×10^{-8} M); ○, *D. melanogaster/EcoR_I* $\geq 20,500$ bp segments (7.16×10^{-8} M).

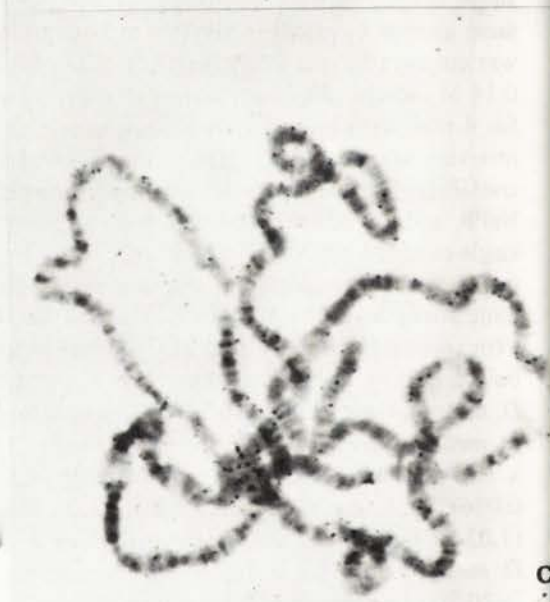
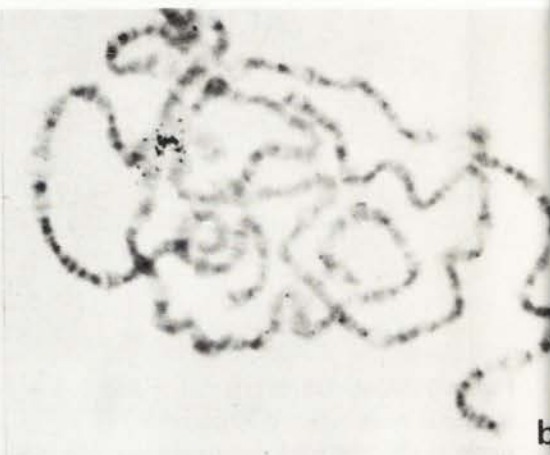


Figure 8. *In situ* hybridization of complementary RNA templated by spared segments. *E. coli* RNA polymerase was used to transcribe spared segments into high specific activity complementary ^3H -RNA. Conditions for RNA synthesis, purification, and *in situ* hybridization to polytene chromosomes were those described by Gall and Pardue (1971). Hybridization was performed for 15 hours at 65°C using $15\ \mu\text{l}$ of a solution containing 7×10^6 cpm/ml of ^3H -RNA (~ 22 Ci/mole), 1 mg/ml of *D. melanogaster* 18 plus 28S rRNA, 0.3 M NaCl, 0.03 M sodium citrate, pH 7.4. The slides were exposed for two weeks. (a) Hind spared segments. (b) Hae spared segments. (c) EcoR_I spared segments.

Interpretation of the fraction of DNA in spared segments It will be recalled that one of the motivations for these experiments was to test the g-region model shown in Figure 1. In accord with expectation from this model, and equation (1), there are some spared long segments. However, when these long segments are characterized further, it turns out that they consist predominately of simple sequence DNAs that have been encountered by others. Thus given the assumptions made, these experiments offer no support at all for this model for the *Drosophila* chromosome. Can the evidence adduced here rule out the g-region model? One way of approaching this question is first to state what fraction of spared segments could be other than the simple sequences, then to put this number, F_0 , into equation (1) and ask how long s must be to accommodate the observations. Since the Hind and Hae spared segments amount to 5% of the DNA, and of these all but 10% can be accounted for as simple sequences, only 0.5% could correspond to spared g-regions. Of the EcoR_I spared DNA, one half can be demonstrated as belonging to the classes of simple sequences leaving a maximum of 5%. When these values are substituted into equation (1), and the error in $1/p$ (Table 1) is included, one may estimate the *minimum* value of s required to be consistent with the model in Figure 1. As shown in Figure 9, the Hae experiments would demand an $s > 1,500$, the Hind $> 3,700$, and EcoR_I $> 4,600$ bp. Clearly the EcoR_I experiments are most demanding. To maintain the g-region model one must require a value of s greater than 5-6,000 bp. This number conflicts *sharply* with the value of s estimated from the folded ring experiments mentioned at the outset.

Another way to approach this problem is to assume that $s = 1,000$ bp, then to rearrange equation (1) and calculate the fraction of the DNA, γ , which could be organized in g-regions with this repeat length. Taking maximum values of non-simple sequence DNA spared to be 5% for EcoR_I and 0.5% for Hind and Hae digests, it can be calculated that γ must be less than 8.2%, 1.8%, and 1.7% to agree with the data obtained with EcoR_I, Hind and Hae respectively. Hence it appears that less than 2% of *D. melanogaster* euchromatic DNA is organized in g-regions with repeat lengths of 1,000 bp or less.

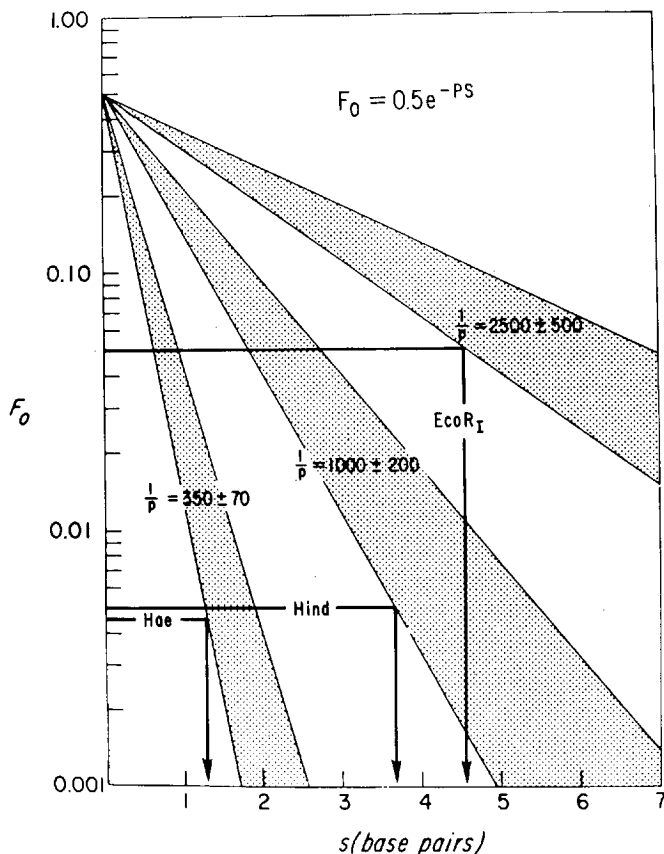


Figure 9. The fraction of DNA in spared g-regions expected assuming the density of recognition sites determined in Figure 4 and shown in Table 1. Tandemly repeating units of length s will contain an average of ps recognition sites, therefore $\exp(-ps)$ will contain none if they are distributed randomly. The maximum fraction of non-simple sequence DNA found in spared long segments is shown by the horizontal lines corresponding to the three different digests. This value defines a *minimum* value of s in order to maintain the g-region model shown in Figure 1. The shaded portions correspond to the uncertainty of p as listed in Table 2. We read this graph as follows: Under the assumption that EcoRI sites are distributed randomly, a g-region model in which $s = 4,600$ bp is consistent with the data; any value of s less than this is not. Similar statements can be made for the other two nucleases.

Before leaving this subject, we should underline some of the basic assumptions in this kind of experiment. First, we assume that the spared segments $\geq 20,500$ bp include all spared g-regions. In actual fact the mean value for g must be near 15,000 bp so that our class of 'spared segments' could only hope to include some portion of the spared g-regions. Secondly, we calculate the

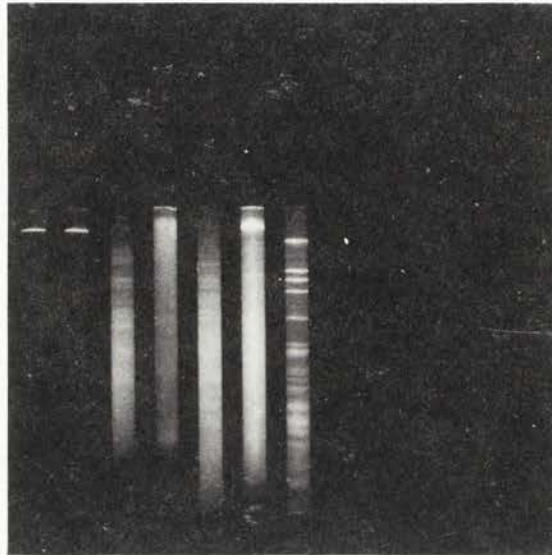
expected fraction of spared g-regions assuming random disposition of endonuclease recognition sites. If there were any preferential disposition (as might result from some biochemical requirement for transcription or translation) then this evidence would be very weak indeed. The evidence to the contrary is that distribution of segment lengths appears random (Figure 4) and the value of $1/p$ is about that expected on a random basis for three different endonucleases. Therefore, we were surprised that we could not find any spared g-regions, if the g-region model had any validity.

Recently Botchan et al. (1974) have studied the cleavage of mouse DNA by endonuclease $EcoR_I$, using techniques similar to those reported here. They found most of the terminally digested mouse DNA in small segments which had a random size distribution with $1/p = 2,250$ bp. The rest of the DNA was found in long segments ($> 6,000$ bp) which were shown to consist almost entirely of rapidly renaturing satellite sequences. Thus, our general conclusions seem applicable to mouse as well as *Drosophila* DNA.

What are the $EcoR_I$ spared segments that are not simple sequence DNAs? As mentioned above, about 50% of the $EcoR_I$ spared segments belong to classes of simple sequences, associate rapidly, and are homologous to the chromocenter, while the remainder are slowly associating, codistribute with main band DNA and are homologous to various parts of the euchromatin. Could these be spared g-regions? If they were, they would represent only 5% of 5,000 such regions or 250 different kinds of sequences. If each of these g-regions contained repeating sequences 1,000 bp long, the total sequence complexity would only be 250,000 bp. As shown in Table 2, 9.4% of *E. coli* DNA is spared by $EcoR_I$ in segments longer than 20,500 bp. Taking 4×10^6 bp for the *E. coli* chromosome, the sequence complexity of the spared segments would be about 376,000 bp - a value comparable to that expected for the $EcoR_I$ spared segments of *Drosophila*. If both kinds of spared DNA were now digested by Hae, one might expect to resolve zones containing individual sequences. An experiment of this kind is shown in Figure 10. $EcoR_I$ spared segments were collected by electrophoresis and slicing. These were rerun on 2% gels to produce the bright zones migrating at the limiting rate shown in the left hand pair of gels (a,b). Both samples were terminally digested with Hae and the resulting electrophoregrams are shown in the next pair of gels (c-*E. coli*, d-*D. melanogaster*). As can be seen, the *E. coli* segments are beginning to resolve themselves into a large number of zones. This is to be expected because we are beginning to reach the limited number of different kinds of unique sequences in the *E. coli* chromosome. In contrast, that portion of the *D. melanogaster* $EcoR_I$ spared DNA

that is broken by Hae is now distributed in a broad blur of more numerous unresolved zones. Thus, the complexity of these segments must be substantially greater than that of the *E. coli* DNA. This again conflicts with the simple g-region model.

The formation of folded rings from EcoR_I segments Since folded rings are formed by annealing fragments that have been partially resected with exonuclease III, or its equivalent, rings will only form if nearly identical sequences are found near both terminals. If the g-region model (Figure 1) is strictly valid, the terminal digestion



(a)(b) (c) (d) (e) (f) (g)

Figure 10. Digestion of *E. coli* and *D. melanogaster* EcoR_I spared segments with endonuclease Hae. *E. coli* and *D. melanogaster* DNA were terminally digested with EcoR_I and electrophoresed through preparative 0.85% agarose gels. The $\geq 20,500$ bp segments were eluted and purified as described in the legend to Figure 6, digested with excess Hae, and electrophoresed through analytical 2% agarose gels. (a) *E. coli* spared segments (0.035 μ g), control incubation. (b) *D. melanogaster* EcoR_I spared segments (0.035 μ g), control incubation. (c) *E. coli* EcoR_I spared segments (0.85 μ g), digested with Hae. (d) *D. melanogaster* EcoR_I spared segments (0.85 μ g), digested with Hae. (e) Total *E. coli* DNA (1.2 μ g), digested with Hae. (f) Total *D. melanogaster* DNA (1.2 μ g), digested with Hae. (g) λ DNA (0.8 μ g), digested with Hae.

products should never be able to form folded rings unless the segment contained an entire g-region. Clearly, if a recognition site is located in every tandem repeat, cleavage will produce segments that are not terminally repetitious and therefore would be unable to cyclize. Those segments cut from the non-repetitious portions of the genome should also be unable to cyclize. The only segments capable of forming folded rings should be the intact g-regions themselves. From the discussion above, the only spared long segments that we have been able to find are derived from various classes of simple sequence DNAs, and these are of the order of 30,000 bp (10 μ) long. Therefore, one may ask, are shorter segments capable of forming rings as would *not* be expected by the g-region model? Terminal EcoR_I digests of *D. melanogaster* DNA and λ DNA were optimally treated with exo III in parallel with shear-broken *D. melanogaster* fragments. After annealing for two hours, electronmicrographs were made, rings scored and contour lengths measured. The results are shown in Figure 11. The λ /EcoR_I segments do not cyclize at all. The folded rings made from shear-broken *D. melanogaster* fragments display the characteristic decrease in ring frequency as the fragment length increases. The EcoR_I segments display an unexpected frequency of rings in the 3,000 bp (1 μ) range. As the segment length increases, one sees first a decrease, then an increase consistent with the finding that long segments are simple sequences. We have no idea of what kind of sequences are forming these small rings, but the observation sharply conflicts with the simple g-region model. Cyclization of restriction segments has also been observed with HeLa DNA (Hardman and Thomas; Hamer and Pyeritz, unpublished observations) and with mouse main-band and guinea pig DNA (Southern and Roizes, 1974).

Some surprising Hae segments In the course of this work we electrophoresed Hae digests of *D. melanogaster* DNA through dilute gels (0.85% agarose) in order to collect spared segments. After fluorochroming a number of prominent, sharp zones may be seen (Figure 12). These zones are observed from a variety of *D. melanogaster* sources including several cell culture lines, embryos and pupae. Some of these segments were resistant to multiple endonucleases, some were broken further. Table 3 shows some preliminary information about these species. Zones B, C and E are produced from purified mitochondrial DNA. Zones D, G and F plus other smaller segments not listed here appear to be derived from species that can be purified by various isopycnic sedimentations. Thus, they may be derived from relatively simple sequence DNAs. The fractional weight and relative number of these segments per haploid genome seem exceptionally high. The origin and chromosomal location of these segments are currently being studied.

TABLE 3
Segments of distinctive length cut by Hae

<i>D. me</i> λ.Hae segment	Length (base pairs)	Percent of total DNA	Copies per haploid cell	Subcellular origin
A	≥20,500	5.0	<366	Nuclear
B	8,700	0.53	92	Mitochondrial
C	5,450	0.44	121	Mitochondrial
D	4,750	0.94	298	Nuclear
E	3,560	0.23	96	Mitochondrial
F	3,000	0.18	89	Nuclear
G	1,900	0.26	209	Nuclear

The lengths of segments in the various zones appearing in Figure 12 were assigned by coelectrophoresis with λ/EcoRI segments and T₇/Hae segments. The relative amounts were measured by the intensity of fluorescent light in a scanning fluorimeter. The number of copies of each segment per haploid cell was calculated, assuming a c-value of 1.7×10^8 bp. The subcellular origins of the various segments were determined by digesting DNA prepared from mitochondria, nuclei and chromatin.

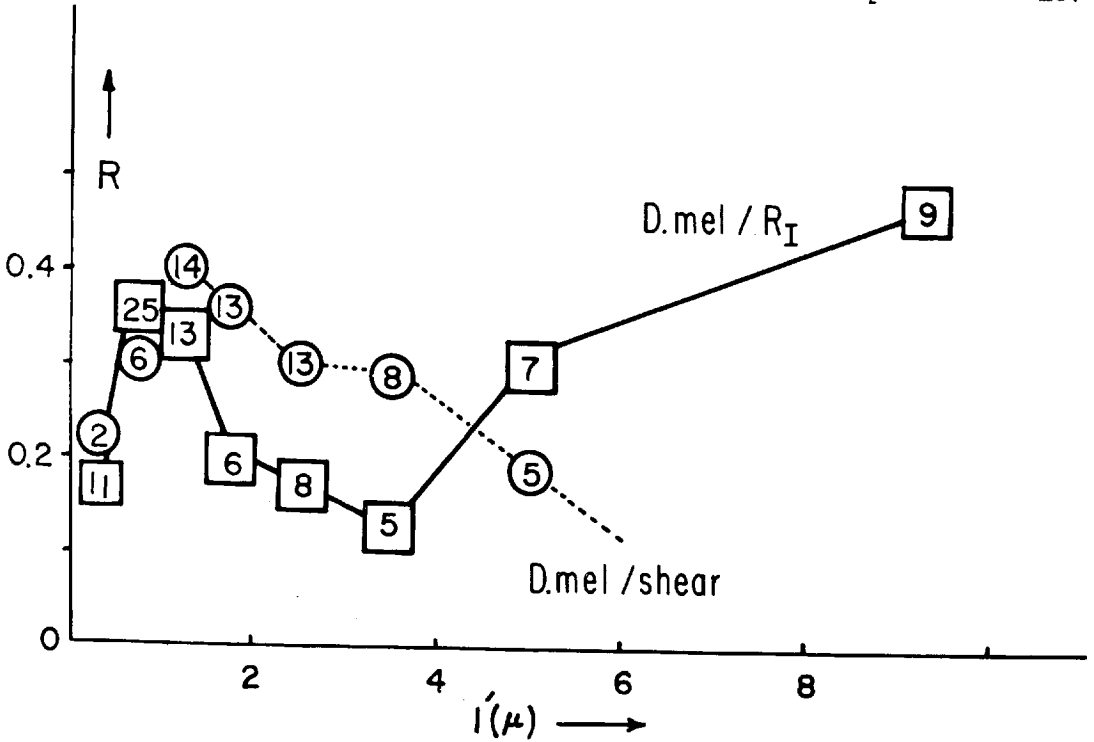


Figure 11. The formation of folded rings from $EcoR_I$ segments of *Drosophila* DNA. Terminal $EcoR_I$ digests were treated with Exo III to remove approximately 500 nucleotides from each terminus brought to 2x SSC, annealed at 60°C for 2 hours, spread by the aqueous technique for EM. For details see Thomas et al. (1973). Fields containing rings were photographed and the contour lengths of rings and linear fragments were measured. The molecules were grouped according to the length of the double helical portion (l') and the fraction of rings plotted against the length class. This method gives a value of R , the fraction of rings, that is about 2x too high due to biased counting of rings. The number of rings seen in any length category is quoted within the square data points. From this value, and the corresponding value of R , the counting error can be estimated. The dotted line and circles show the results of a parallel experiment with shear broken fragments.

CONCLUSIONS

The experiments described here were designed to test the g-region model for the organization of nucleotide sequences in *Drosophila* DNA. In each case the model was found wanting. How this conflict will be resolved is still not completely clear, but the general direction of future experiments is evident. The application of restriction endonucleases and subsequent isolation and characterisation of unique chromosomal segments is now a realizable objective.

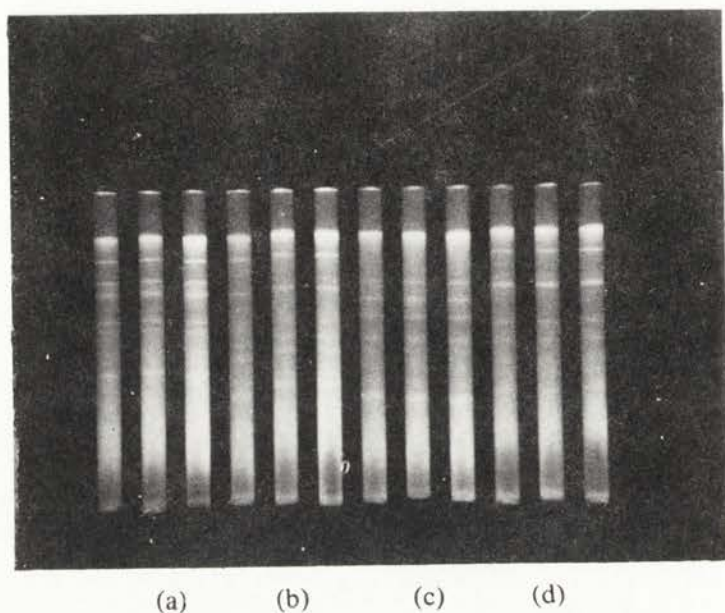


Figure 12. Segments of distinctive length cut by endonuclease Hae. *D. melanogaster* DNA was digested with Hae or with Hae plus another endonuclease and electrophoresed through 0.85% agarose gels as described in the legend to Figure 5. *Haemophilus parainfluenzae* endonuclease HpaII was a gift from Dr H. Kopecka. Each reaction was done using three different amounts of DNA: (from left to right) 1.8 μg , 2.7 μg and 3.6 μg . (a) Hae digests. (b) Hae plus EcoRI digests. (c) Hae plus Hind digests. (d) Hae plus HpaII digests.

ACKNOWLEDGMENTS

We thank W. Gilbert for endonuclease Hind, H. Kopecka for endonuclease HpaII, and G. Fareed for SV40 DNA. This work was supported by the National Science Foundation (GB-3118X).

REFERENCES

- Botchan, M., G. McKenna and P.A. Sharp. 1974. Cleavage of mouse DNA by a restriction enzyme as a clue to the arrangement of genes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 383.
- Danna, K.J., G.H. Sack and D. Nathans. 1973. Studies of SV40 DNA. VII. A cleavage map of the SV40 genome. *J. Mol. Biol.* 78: 363.
- Gall, J.G. and M.L. Pardue. 1971. Nucleic acid hybridization in cytological preparations. In *Methods in Enzymology XII C* (L. Grossman and K. Moldave, ed.). Academic Press, New York.
- Middleton, J.H., M.H. Edgell and C.A. Hutchison III. 1972. Specific fragments of ϕ X174 deoxyribonucleic acid produced by a restriction enzyme from *Haemophilus aegyptius*, endonuclease Z. *J. Virol.* 10: 42.
- Montroll, E. and R. Simha. 1940. Theory of depolymerization of long chain molecules. *J. Chem. Phys.* 8: 721.
- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 405.
- Schneider, I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morph.* 27: 353.
- Smith, H.O. and K.W. Wilcox. 1970. A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties. *J. Mol. Biol.* 51: 379.
- Southern, E.M. and G. Roizes. 1974. The action of a restriction endonuclease on higher organism DNA. *Cold Spring Harbor Symp. Quant. Biol.* 38: 429.
- Thomas, C.A. Jr, R.E. Pyeritz, D.A. Wilson, B.M. Dancis, C.S. Lee, M.D. Bick, H.L. Huang and B.H. Zimm. 1974. Cyclodromes and palindromes in chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 353.
- Yoshimori, R.N. 1971. A genetic and biochemical analysis of the restriction and modification of DNA by resistance transfer factors. Ph.D. thesis, University of California, San Francisco Medical Center.

Circular DNA in the yeast *Saccharomyces cerevisiae* in relation to informosomal DNA and mitochondrial genetics

G. D. Clark-Walker and George L. Gabor Miklos

The yeast *Saccharomyces cerevisiae* contains two major types of circular DNA with buoyant densities of 1.701 and 1.684 gm/cm³. Since the 1.701 circular DNA has the same density as chromosomal DNA we examined whether it has a nuclear location by using the small minicircular mitochondrial DNA in a *petite* mutant as a marker of cytoplasmic contamination of nuclei. We find that the 1.701 circular DNA is absent from purified nuclei, and this argues strongly against a possible informosomal role for this DNA. Secondly we have inquired into possible mechanisms for generating the small minicircular mitochondrial DNA from larger circular molecules. These studies have led us to examine some of the complexities in mitochondrial genetic systems in yeast in terms of insertion and excision events between circular molecules.

One of the important observations in relation to the organization of DNA in eukaryote chromosomes is that a percentage of this DNA has the ability to form circles after resectioning with an exonuclease (Thomas et al., 1970). Furthermore ribosomal DNA can, under certain circumstances, and in some organisms, be amplified, whereupon it is found in the form of extra-chromosomal circles (Hourcade et al., 1973). This fact may reflect the general ability of chromosomal DNA to circularize in particular cell types or it may only be a property unique to ribosomal cistrons. Clearly, however, the finding of circular DNA both *in vivo* and its production *in vitro* is of general interest, not only because it may give insight into the organization of DNA in chromosomes, but because we feel it is relevant to the structural peculiarities of mitochondrial DNA in yeast.

In this context, we have examined possible ways in which small circular DNA molecules are formed from larger DNA. Specifically we have asked if circular DNA molecules of nuclear buoyant density are produced from yeast chromosomes and also how the observed small circular DNA in certain *petite* mutants is produced from the larger mitochondrial DNA.

CIRCULAR DNA IN YEAST

Yeast is an eukaryote microorganism and normally contains at least two major types of circular DNA which differ both in buoyant

density and size range. One class of circular DNA, termed ρ DNA, has a buoyant density of 1.684 gm/cm^3 , a length of 21-25 μm and is thought to be localised in the mitochondrion (Hollenberg et al., 1970; Petes et al., 1973). The second major class of circular DNA, termed σ DNA, has a buoyant density of 1.701 gm/cm^3 , which is similar to the density of main band DNA. This circular DNA has two distinct size classes of 1.9 and 2.8 μm and can be found in the cytoplasm (Clark-Walker, 1972, 1973), but because of its buoyant density it may have a chromosomal origin and exist in the nucleus as well.

In this communication we examine some of the phenomena associated with both these major types of circular DNA. We considered σ DNA a potential candidate for informosomal DNA because of its nuclear buoyant density. In other organisms, it has been suggested that informosomal DNA is produced in the nucleus by gene amplification and subsequently transported to the cytoplasm (Bell, 1971). Therefore we have examined whether σ DNA occurs in the nucleus by using as a marker of cytoplasmic contamination of nuclei the minicircular mitochondrial DNA in a *petite*. On the other hand, in considering ρ DNA, we start with the observation that some *petite* mutants contain a minicircular ρ DNA varying in size from 0.05 to 1.5 μm (Hollenberg et al., 1972a; Clark-Walker and Miklos, 1974a). On the basis of this finding we have developed a hypothesis which we feel provides a new insight into the phenomena of mitochondrial recombination and polarity. This hypothesis also suggests a mechanism for the generation of *petites*.

Circular DNA of chromosomal buoyant density

The first major type of circular DNA found in the yeast cell consists of a population of covalently closed molecules of buoyant density 1.701 gm/cm^3 (with a small shoulder on the heavy side of the 1.701 peak) and a major size class of 1.9 μm . Three minor size classes are also consistently observed and these have lengths of 2.8, 3.9 and 5.8 μm . In the strain examined in detail, these four classes occur with frequencies of 74%, 4%, 16% and 6% on a number basis and are collectively termed σ DNA.

In contrast to circular ρ DNA which is known to be located in mitochondria, the location and origin of σ DNA has not been definitely determined. A claim has been made that the σ DNA has a nuclear location (Stevens and Moustacchi, 1971) but it is also known to be recovered from a light particulate fraction in sucrose gradients (Clark-Walker, 1972). These facts suggested to us that it may have a nuclear origin, followed by subsequent transport to a cytoplasmic location, thus fulfilling one of the criteria invoked for informosomal DNA. Two further facts seemingly support this suggestion; σ DNA has the same buoyant density as main band DNA, and in contrast

to ρ DNA it is not eliminated by growth limiting concentrations of the dye ethidium bromide. Therefore, we have investigated whether ρ DNA has a nuclear location because, if this were so, it would suggest a possible function for ρ DNA as a gene amplification product.

Circular DNA in isolated nuclei In order to demonstrate that we have in fact achieved a significant purification of nuclei, we have used the minicircular mitochondrial DNA from one of our *petite* strains as a cytoplasmic marker. This minicircular ρ DNA was extremely useful, since its size distribution, 0.05-1.6 μ m, is less than that of the smallest ρ DNA class (1.8-2.1 μ m), and hence ρ DNA could always be reliably determined using the electron microscope (Figure 1a).

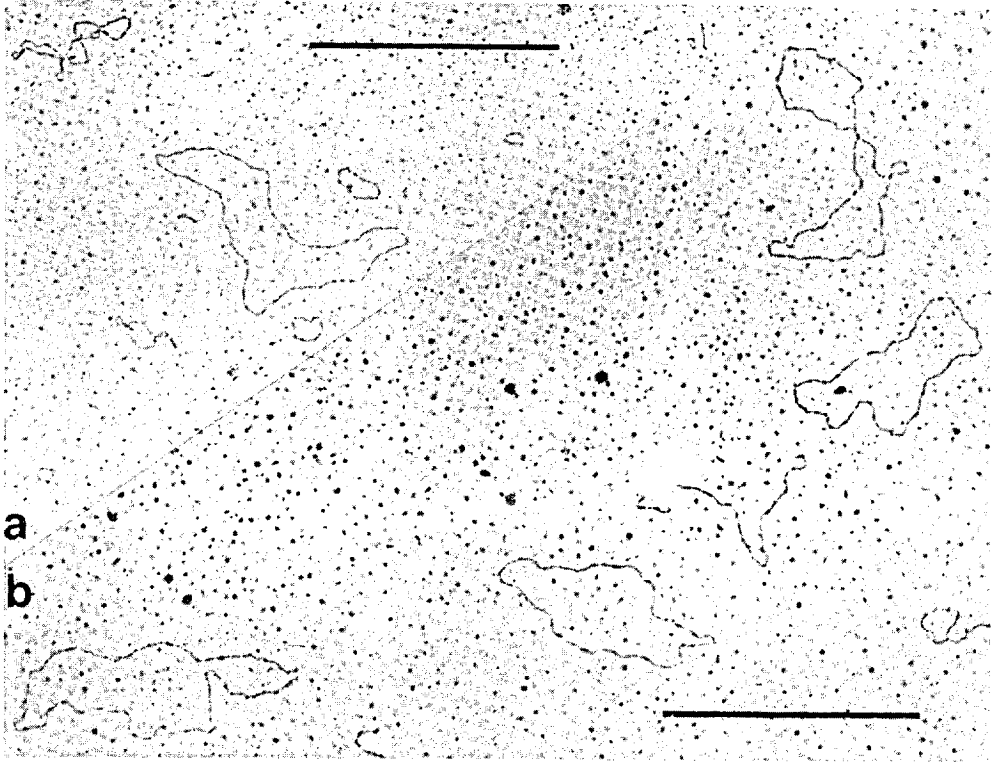


Figure 1. Circular DNA molecules from a cytoplasmic particulate fraction showing: (a) minicircular ρ DNA forms from a *petite* mutant retaining ρ DNA of buoyant density 1.676 g/cm³ together with a 2.8 μ m open circular ρ DNA molecule of buoyant density 1.704 g/cm³ and a 1.9 μ m supercoiled ρ DNA molecule of buoyant density 1.701 g/cm³, (b) molecules of nuclear buoyant density showing the 1.9 and 2.8 μ m size classes. The bar in each case represents 1 μ m.

The details of growth conditions, radioactive labelling, nuclei isolation and DNA ultracentrifugation have been described previously (Clark-Walker and Miklos, 1974a).

We have found that by using minicircular ρ DNA as a monitor of the purity of nuclei, this DNA together with circular ϕ DNA is coordinately removed during nuclear purification (Table 1). This strongly suggests that yeast nuclei do not contain ϕ DNA and that the residual ϕ DNA observed in isolated yeast nuclei is due to a small remaining cytoplasmic contamination. These results indicate that ϕ DNA does not normally have a nuclear location, but they do not necessarily rule out the possibility that ϕ DNA has a chromosomal origin if movement from the nucleus is extremely rapid.

TABLE 1

Frequencies per haploid genome (5000 μ m) of circular and linear molecules observed in whole cell and purified nuclear preparations in two experiments

	Number of molecules	
	Minicircular ρ DNA	ϕ DNA
Whole homogenate 1	72	72
Whole homogenate 2	89	53
Nuclei preparation 1	10	10
Nuclei preparation 2	7	2

Although the location of the circular DNA would suggest possible functions, an alternative approach may be obtained from correlations with specific features of the DNA such as buoyant density. Another facet to this problem is that there are at least two unique size classes of circular molecules in ϕ DNA with lengths of 1.9 and 2.8 μ m (Figure 1b). It is from buoyant density analyses that we have obtained a possible clue to the function of the 2.8 μ m size class.

A possible role for the 2.8 μ m size class DNA Even though the 2.8 and 5.8 μ m size classes occur in equal frequencies, we consider that the latter is likely to represent a dimer of the 2.8 μ m form. A suggested function for these two size classes has come from studies with a spontaneous mutant strain in a *petite* completely lacking ρ DNA.

When circular DNA is prepared from a strain resistant to the toxic effects of copper ions, there is an increase in the heavy side of the 1.701 gm/cm^3 buoyant density peak as seen in the analytical ultracentrifuge. This shoulder has an approximate density of 1.704 gm/cm^3 and its area increases from 9% in the parent to 19% of the total area in the mutant.

Examination of the circular DNA from this strain in the electron microscope reveals that the 2.8 and 5.8 μm forms have increased from 10% to 22%. In addition, the 3.8 μm size class has decreased in frequency from 16% to 6% in this analysis (Table 2).

TABLE 2

Frequencies of circular molecules observed in the copper resistant and parental strains

Strain	Size classes (μm)				No. of molecules
	1.9	2.8	3.8	5.8	
5ep^{-1}	74%	4%	16%	6%	165
5ep^{-1} Cu resistant	72%	10%	6%	12%	158

The correlation between the increase in the 1.704 gm/cm^3 shoulder, and the increase in frequency of the 2.8 and 5.8 μm size class of molecules, suggests that the buoyant density of these forms is indeed 1.704 gm/cm^3 . It is interesting to note that ribosomal DNA cistrons in yeast also have a buoyant density of 1.704 gm/cm^3 (Kaback et al., 1973). Moreover the length of the cytoplasmically located 2.8 μm size class is sufficient to code for the ribosomal RNA precursor molecule; however, as discussed previously, that of the 1.9 μm molecules is not (Clark-Walker and Miklos, 1974a). Thus the cytoplasmic location of the 2.8 μm class raises a number of questions, if indeed the DNA is ribosomal. Either it is produced in the nucleus and rapidly exported, or it is not in fact ribosomal DNA and its density of 1.704 is a fortuitous coincidence.

Finally, we wish to emphasise that there are at least two unique components of oDNA and these components may not be related in their function, nor necessarily in their site of origin.

Properties of circular mitochondrial DNA

In the foregoing experiments we utilised the presence of mini-circular mitochondrial DNA molecules in a *petite* mutant without

inquiring into the problem of how these small circles are formed. However, we feel that the production of these circles is of considerable interest because it may provide an insight into those properties of DNA that are necessary to generate circles from larger DNA molecules. We have previously discussed some of the structural peculiarities of yeast mitochondrial DNA which we think facilitate excision events from the circular ρ DNA genome, and have developed a hypothesis to account for the generation of the observed small circular DNA in *petites* (Clark-Walker and Miklos, 1974b).

One of the properties of ρ DNA which may facilitate elevated levels of exchange, resulting in the formation of smaller incomplete genomes, may be sites of partial base sequence homology. Such sites may be AT rich regions since it is known that ρ DNA is extraordinarily rich in AT base pairs (82% AT) and that AT rich regions are interspersed throughout the mitochondrial genome (Piperno et al., 1972).

An additional finding in some *petites* is that the smaller residual ρ DNA genome can become tandemly repetitious resulting in very large molecules (Hollenberg et al., 1972b). For instance Van Kreijl et al. (1972) have shown for one *petite* that the mitochondrial DNA consists of large molecules which occur as a perfect repetition of a sequence 300 base pairs in length. One way we feel this can be brought about is by multiple cycles of insertion events between smaller ρ DNA molecules.

A consideration of the production of these large tandemly repeated sequences by insertion of smaller circles has led us to inquire into the general consequences of insertion and excision cycles in normal circular ρ DNA molecules. The processes we discuss below may offer some novel ways in which phenomena associated with the mitochondrial recombination system in yeast can be re-evaluated.

Mitochondrial recombination The discovery of the cytoplasmic inheritance and mitochondrial DNA location of antibiotic resistance genes enabled the development of a formal mitochondrial genetic system in yeast (Thomas and Wilkie, 1968; Coen et al., 1970; Linnane et al., 1968). Two main phenomena are observed in mitochondrial genetic studies. In some crosses mitochondrial parental types are not recovered in equal frequency and this is termed parental polarity. Additionally, mitochondrial reciprocal recombinant types are not obtained in equal frequency in some crosses and this is termed reciprocal recombinant polarity.

In evaluating these phenomena three sets of complexities need to be considered, these being problems due to multiple molecule populations, the inherent properties associated with mitochondria in

yeast, and the circular nature of the mitochondrial genome. The first general consideration is that involving multiple molecule populations, as it is known that haploid yeast cells contain between 20 and 40 ρ DNA molecules (Nagley and Linnane, 1972). Since mating involves cell fusion, it may be expected that more than one ρ DNA from each mating cell is involved in recombination. Accordingly, we believe that the following variables need to be considered in discussing interactions between multiple molecule populations.

- (a) Different haploid strains contain different numbers of ρ DNA molecules (Williamson, 1970). If all molecules interact at random in a cross between two strains which have different amounts of ρ DNA, the frequency of parental types from the cross will be directly related to the input ratio of the parental ρ genomes.
- (b) The time available for interactions may vary from crosses of different strains, and the rate of recombination of ρ DNA will affect the number of rounds of interactions occurring between molecules. The number of rounds will also influence the total amount of recombination in a cross.
- (c) The frequency of dimers may vary from strain to strain, and recombination of a dimer with a monomer can produce different results from monomer-monomer interactions. A significant consideration here is the dominance relationships of the particular genes involved.

The second consideration concerns properties peculiar to the yeast system itself.

- (a) Not all ρ DNA molecules in a mating form may interact. The 'effective pool size' may vary because only those ρ DNA molecules adjacent to the fusion site may have the opportunity to recombine. Buds arising from the fusion bridge may contain recombinant ρ DNA molecules, whereas those arising from distal regions of the cell may not be recombinant. These constraints must affect the total amount of recombination in a cross.
- (b) Nuclear genes, as well as influencing the above processes, may also cause a polarity of parental types. For example, mating type appears to be important in the transmission of mitochondrial genes to buds arising from the zygote (Callen, 1974).

The foregoing considerations apply regardless of whether linear or circular molecules are considered. However, further complexities

occur if interactions take place between circular ρ DNA molecules. Some of the difficulties in dealing with circular genetic maps have previously been considered in evaluating recombination in bacteriophage systems (Foss and Stahl, 1963; Stahl, 1967), but as we shall see below, our mechanism differs in general from the phage ones, although some of the complexities pertain to both.

It is our contention that mitochondrial recombination occurs by insertion between two circular genomes to form a dimeric circular molecule, and this is followed by excision to produce two complete monomeric circular genomes. Another possibility resulting from a second exchange event following initial insertion is the formation of a catenane rather than two monomers. Catenane formation temporarily removes recessive genes from the analysis. Furthermore in the simplest case of recombination, both insertion and excision are thought to take place by homologous pairing and this can occur anywhere in the genome. With multiple molecule interactions between two ρ DNA populations, such as ABC and abc, half of the insertions on a random basis are ABC/ABC and abc/abc ones, and it is only the remaining half, ABC/abc, which can produce recombinant phenotypes. Some specific examples of the consequences of recombination between circular molecules have been discussed in detail previously (Clark-Walker and Miklos, 1974b). In the next section we specifically examine the factors affecting parental and recombinant polarity.

Parental polarity Parental polarity in mitochondrial genetic studies will result from crosses where the haploid strains contain different numbers of ρ DNA molecules. Of significance in this situation will be the mating type of the strains involved, and the specific genetic influences controlling the site of budding.

Additionally, transient parental polarity may result from the time available for recombination. If the time available is very short, so that most molecular interactions have produced dimers but only a few have undergone excision at the time of analysis, then very extreme polarity of either parentals or recombinants can be found because most molecules still exist in a dimeric form and the recessive alleles in the dimer are covered by the dominant ones. There can be low or very high frequencies of recombination, depending on the dominance-recessive relationships of the genes involved in the cross.

The previous considerations have only involved monomer-monomer interactions, but most mitochondrial DNA populations contain a significant proportion of dimers which presumably are in some form of equilibrium with monomeric forms. For example, in a cross in which each strain has 10% of its ρ DNA as dimers, and interactions are at random, 18% of the heterogenomic interactions produce trimers. In

the simplest case of a trimer breaking down to produce a dimer and a monomer, recessive genes in a dimer will again be excluded from phenotypic expression by their dominant alleles on the same molecule. If the dimer subsequently undergoes excision, the recessive gene may achieve phenotypic expression.

The foregoing discussions have dealt with the situation where insertion and excision reactions occur in homologous regions, but as we have indicated previously in formulating mechanisms for the generation of incomplete genomes, certain special properties of ρ DNA may lead to insertion by an alternative process to homologous pairing. Thus insertion and excision reactions may involve multiple sites of AT rich partial base sequence homology leading to complex dimeric forms of altered gene order. If such is the case then not only parental polarity but reciprocal recombinant polarity can be generated.

Reciprocal recombinant polarity Reciprocal recombinant polarity amongst ρ DNA linked genes was first described by Coen et al. (1970) and numerous studies have confirmed these initial observations. We consider that there are at least four possible ways in which this phenomenon can be generated by insertion between circular ρ DNA molecules one of which contains any of the following structural rearrangements.

1. Duplication of a sequence
2. Transposition of a sequence
3. Deletion of a non-essential region of partial base sequence homology
4. Insertion at sites of AT rich partial base sequence homology which produce dimers that are 'out of register' (transposition dimers).

In all of these cases a common feature is the generation of defective and partially duplicated genomes following, in the simplest instance, excision at homologous regions. Moreover, the phenotypic expression of the genes located in the reciprocal partially duplicated molecules will depend on the dominance-recessive relationships involved. Further breakdown of the partially duplicated molecules to yield complete monomers will eliminate one or other of the duplicated regions by preferentially removing genes in this region into *petites*.

Thus in a cross in which the ρ DNA of one strain contains one of the structural rearrangements mentioned, there should be a high frequency of defective molecules lacking a complete ρ DNA genome. The

incomplete genomes can give rise to *petites* by segregation into buds. Owing to the methods of experimental analysis employed in the past the genetic content of *petites* has not been assayed and this can produce a loss of genes from the cross. Thus in order to recover all input genetic markers from a cross showing reciprocal recombinant polarity, both ρ^+ and ρ^- cells need to be analysed. Furthermore, because of subsequent ρ DNA information decay in *petites*, these mutants need to be analysed as soon as buds are produced following zygote formation. Therefore a prediction is that in a cross showing highly polar reciprocal recombination there will be an increased frequency of *petite* production at least during the time immediately following zygote formation. This prediction is only applicable if recombinant polarity is due to ρ DNA structural rearrangements between various yeast strains. Structural rearrangements of the type discussed are suggested by the results of Howell et al. (1973) who find that varying polarity values are obtained in crosses of different strains to a common tester strain. It should also be pointed out that, as with many other processes that are under some form of nuclear control, polarity may well be influenced in some cases by direct nuclear intervention and in these cases an increase in *petite* formation is not a necessary concomitant.

Finally suggestive evidence that structural rearrangements may occur in ρ DNA of ρ^+ cells comes from the finding of repeats and tandem inverted repeats in certain *petites* (Locker et al., 1974). In relation to this last point concerning the occurrence of inverted regions in ρ DNA we would like to indicate that inversions, although they may lead to an increase in the frequency of *petites*, do not yield reciprocal recombinant polarity.

REFERENCES

- Bell, E. 1971. Information transfer between nucleus and cytoplasm during differentiation. *Symp. Soc. Expt. Biol.* 25: 127.
- Callen, D. 1974. The effect of mating type on the polarity of mitochondrial gene transmission in *Saccharomyces cerevisiae*. *Molec. Gen. Genet.* 128: 321.
- Clark-Walker, G.D. 1972. Isolation of circular DNA from a mitochondrial fraction from yeast. *Proc. Nat. Acad. Sci. USA* 69: 388.
- _____, 1973. Size distribution of circular DNA from petite-mutant yeast lacking ρ DNA. *Eur. J. Biochem.* 32: 263.
- Clark-Walker, G.D. and G.L.G. Miklos. 1974a. Localization and quantification of circular DNA in yeast. *Eur. J. Biochem.* 41: 359.
- _____, _____. 1974b. Mitochondrial genetics, circular DNA and the mechanism of the *petite* mutation in yeast. *Genet. Res.* In press.
- Coen, D., J. Deutsch, P. Netter, E. Petrochilo and P.P. Slonimski. 1970. Mitochondrial genetics I. Methodology and phenomenology. *Symp. Soc. Expt. Biol.* 24: 449.

- Foss, H.M. and F.W. Stahl. 1963. Circularity of the genetic map of bacteriophage T4. *Genetics* 48: 1659.
- Hollenberg, C.P., P. Borst and E.F.J. Van Bruggen. 1970. Mitochondrial DNA. V. 25 micron closed circular duplex DNA molecules in wild-type yeast mitochondria. Structure and genetic complexity. *Biochim. Biophys. Acta* 209: 1.
- _____, _____, _____. 1972a. Mitochondrial DNA from cytoplasmic *petite* mutants of yeast. *Biochim. Biophys. Acta* 277: 35.
- Hollenberg, C.P., P. Borst, R.A. Flavell, C.F. Van Kreijl, E.F.J. Van Bruggen and A.C. Arnberg. 1972b. The unusual properties of mtDNA from a 'low-density' *petite* mutant of yeast. *Biochim. Biophys. Acta* 277: 44.
- Hourcade, D., D. Dressler and J. Wolfson. 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. *Proc. Nat. Acad. Sci. USA* 70: 2926.
- Howell, N., M.K. Trembath, A.W. Linnane and H.B. Lukins. 1973. Biogenesis of mitochondria. 30. Analysis of polarity of mitochondrial gene recombination and transmission. *Molec. Gen. Genet.* 122: 37.
- Kaback, D.B., M.M. Bhargava and H.O. Halvorson. 1973. Location and arrangement of genes coding for ribosomal RNA in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 79: 735.
- Linnane, A.W., G.W. Saunders, E.B. Gingold and H.B. Lukins. 1968. The biogenesis of mitochondria. V. Cytoplasmic inheritance of erythromycin resistance in *Saccharomyces cerevisiae*. *Proc. Nat. Sci. USA* 59: 903.
- Locker, J., M. Rabinowitz and G.S. Getz. 1974. Tandem inverted repeats in mitochondrial DNA of *petite* mutants of *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci. USA* 71: 1366.
- Nagley, P. and A.W. Linnane. 1972. Biogenesis of mitochondria. XXI. Studies on the nature of the mitochondrial genome in yeast: the degenerative effects of ethidium bromide on mitochondrial genetic information in a respiratory competent strain. *J. Mol. Biol.* 66: 181.
- Petes, T.D., B. Byers and W.L. Fangman. 1973. Size and structure of yeast chromosomal DNA. *Proc. Nat. Acad. Sci. USA* 70: 3072.
- Piperno, G., G. Fonty and G. Bernardi. 1972. The mitochondrial genome of wild-type yeast cells. II. Investigations on the compositional heterogeneity of mitochondrial DNA. *J. Mol. Biol.* 65: 191.
- Stahl, F.W. 1967. Circular genetic maps. *J. Cell. Physiol.* 70, Suppl. 1: 1.
- Stevens, B.J., and E. Moustacchi. 1971. ADN satellite γ et molecules circulaires torsadees de petite taille chez la levure *Saccharomyces cerevisiae*. *Exp. Cell. Res.* 64: 259.

- Thomas, C.A., B.A. Hamkalo, D.N. Misra and C.S. Lee. 1970. Cyclization of eucaryotic deoxyribonucleic acid fragments. *J. Mol. Biol.* 51: 621.
- Thomas, D.Y., and D. Wilkie. 1968. Recombination of mitochondrial drug resistance factors in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Comm.* 30: 368.
- Van Kreijl, C.F., P. Borst, R.A. Flavell and C.P. Hollenberg. 1972. Pyrimidine tract analysis of mtDNA from a 'low-density' petite mutant of yeast. *Biochim. Biophys. Acta* 277: 61.
- Williamson, D.H. 1970. The effect of environmental and genetic factors on the replication of mitochondrial DNA in yeast. *Symp. Soc. Exp. Biol.* 24: 247.

Chromosome structure: evidence against the unineme hypothesis

Judith H. Ford

Cultured cells from human amniotic fluid are sensitive to changes in pH. Exposure of cells to pH values between 7.4 and 7.6 for short periods leads to endomitosis and polyploidy. Cultures exposed to raised pH for half an hour and harvested 4 hours after replacement in 5% CO₂ showed a peak value of 41% endomitoses. To examine whether these cells were at S or G₂ phase at time of exposure to raised pH, tritiated thymidine was added to cultures at the time of treatment. Endomitotic cells showed no label and were thus at G₂ when treated. Further, staining of endomitoses with Feulgen and absorbance measurement at 5600 Å with an integrating microdensitometer showed that the endomitoses had the same DNA content as 'normal' metaphases on the same slide. These findings show that bichromatid chromosomes can split to form two bichromatid chromosomes without further DNA synthesis: hence mitotic chromosomes must normally have at least a bineme structure.

Although DNA is known to be a major component of the eukaryote chromosome, the nature of its organization within the chromosome has not been resolved. Not only is it unclear how the DNA is arranged longitudinally within the chromosome, it is also disputed whether the anaphase chromosome is composed in transverse section of a single DNA helix (unineme), two helices (bineme) or many helices (polyneme). Whilst studies with the electron and light microscopes have tended to support polynemy (Kaufmann et al., 1960; Trosko and Wolff, 1965; Bajer, 1965; Ris, 1966), evidence from studies of replication (Taylor et al., 1957), recombination (Taylor, 1963) and DNAase kinetics (Gall, 1963) have supported a unineme concept.

In this study, experimental results will be presented which demonstrate that human mitotic chromatids are, as a minimum, bineme structures.

MATERIALS AND METHODS

Culture Cells from amniotic fluid specimens (16-30 weeks gestation) were pelleted by centrifugation, suspended in 15% foetal calf serum/85% RPMI 1640 medium in Leighton tubes and cultured in a 5% CO₂/air mixture at 37°C for 8-10 days. When suitable growth was obtained

the cells were subcultured onto coverslips or slides in Petri dishes and prepared for karyotyping after 24-48 hours.

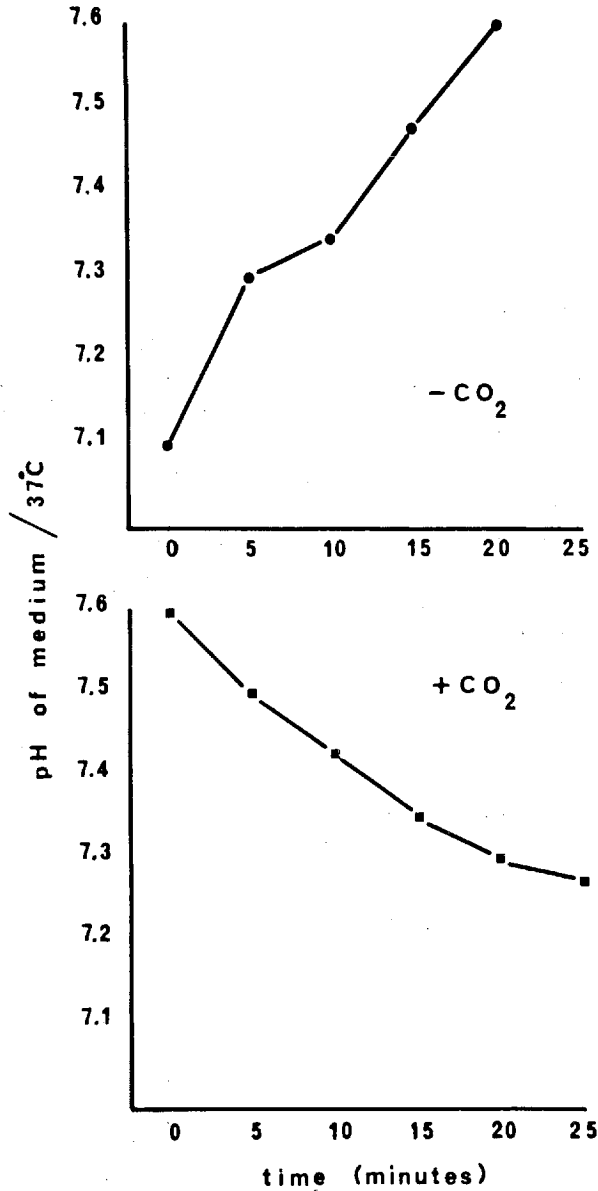


Figure 1. Response of medium RPMI 1640 (in a Petri dish) to changes in CO₂ concentration at 37°C. pH measurements were made by reference to indicator solutions of known pH. (i) Increase in pH observed at various times after the medium is placed in air at 37°C. (ii) Decrease in pH with time after being placed in 5% CO₂ at 37°C.

Raised pH treatment To expose the cells to raised pH, subcultures were removed from the CO₂ environment and placed for periods of 20 min to half an hour in a 37°C incubator where the pH rose to 7.5 (Figure 1). After this time, they were replaced in the 5% CO₂ environment and harvested at various times.

Autoradiography Cells were cultured in medium containing tritiated thymidine (specific activity 8.3 mCi/mg, concentration 1 µCi/ml) and harvested in the usual way. Slides or coverslips were dipped in Ilford K5 or Kodak NTB2 emulsion. After 3-5 days at 4°C, they were developed and stained with Giemsa.

DNA estimations Coverslips were stained in Feulgen (de Tomasi) for 30 min after 8 min hydrolysis in 1N HCl at 60°C. They were then washed in three changes of freshly prepared SO₂ water. Preparations of unstimulated and PHA stimulated lymphocytes were made concurrently. DNA estimations were made on single cells using a Barr and Stroud integrating microdensitometer.

RESULTS

When actively growing cells were exposed to raised pH at 37°C, at various times before harvesting, the proportion of 'normal' diploid metaphases was found to vary (Table 1). The proportion of diploid cells was depressed at two stages. Firstly, between 4 and 6 hours, the frequency of 'endoreduplicated' metaphases (Figure 2) rose to between 35 and 41%. Secondly, after about 17 hours, the frequency of polyploids rose (Figure 2). These two observations will be discussed separately.

Endoreduplication

The observation that the maximum number of endoreduplicated metaphases is present 4 hours after the cultures are exposed to raised pH (Figure 3), suggests that the cells affected by the treatment are at either late S or early G₂ at the time of treatment. To decide between these alternatives, tritiated thymidine was added to the cells 10 min before exposure to raised pH and allowed to remain in the cultures. When the cells were harvested 4 hours after being replaced in the 5% CO₂ environment and observed by autoradiography, many of the diploid metaphases were labelled but endoreduplicated metaphases showed no label or occasionally, a very slight late labelling pattern (Table 2). Thus the cells which were 'endoreduplicated' at metaphase were at the S/G₂ transition at the time of treatment.

If no DNA synthesis occurred between the time of the stimulus to form diplochromosomes and the observation of diplochromosomes, as indicated by labelling experiments, then the diplochromosomes

TABLE 1
Effect of raised pH on metaphase cells

No. of cells	% cells			
	Diploid	Endoredup. diploid	Polyloid	Endoredup. polyloid
<u>Control</u>				
(i) 302	91.5	3.3	4.8	0.3
(ii) 60	100.00	-	-	-
(iii) 37	89.2	5.4	5.4	-
(iv) 194	92.26	2.1	5.15	0.5
<u>pH + 2 hours</u>				
(v) 66	98.48	-	1.52	-
<u>pH + 4 hours</u>				
(vi) 87	58.6	41.4	-	-
(vii) 68	60.3	39.7	-	-
*(viii) 248	59.0	34.9	5.2	0.9
<u>pH + 6 hours</u>				
*(viii) 274	85.4	14.6	4.7	1.1
<u>pH + 8 hours</u>				
(ix) 71	95.8	1.4	2.8	-
<u>pH + 17½ hours</u>				
(x) 22	77.27	4.55	9.10	9.10
<u>pH + 24 hours</u>				
(xi) 78	41.09	6.25	38.46	13.3
<u>pH + 48 hours</u>				
(xii) 53	69.8	5.6	22.4	1.9
(xiii) 26	76.92	-	23.0	-

* Cultures from same patient. Other cultures from different patients.

must have been formed by the splitting of a previously double structure. If this is so, the DNA content of a metaphase cell with 46 diplochromosomes must be the same as one with 46 chromosomes. When slides were stained with Feulgen and the DNA content

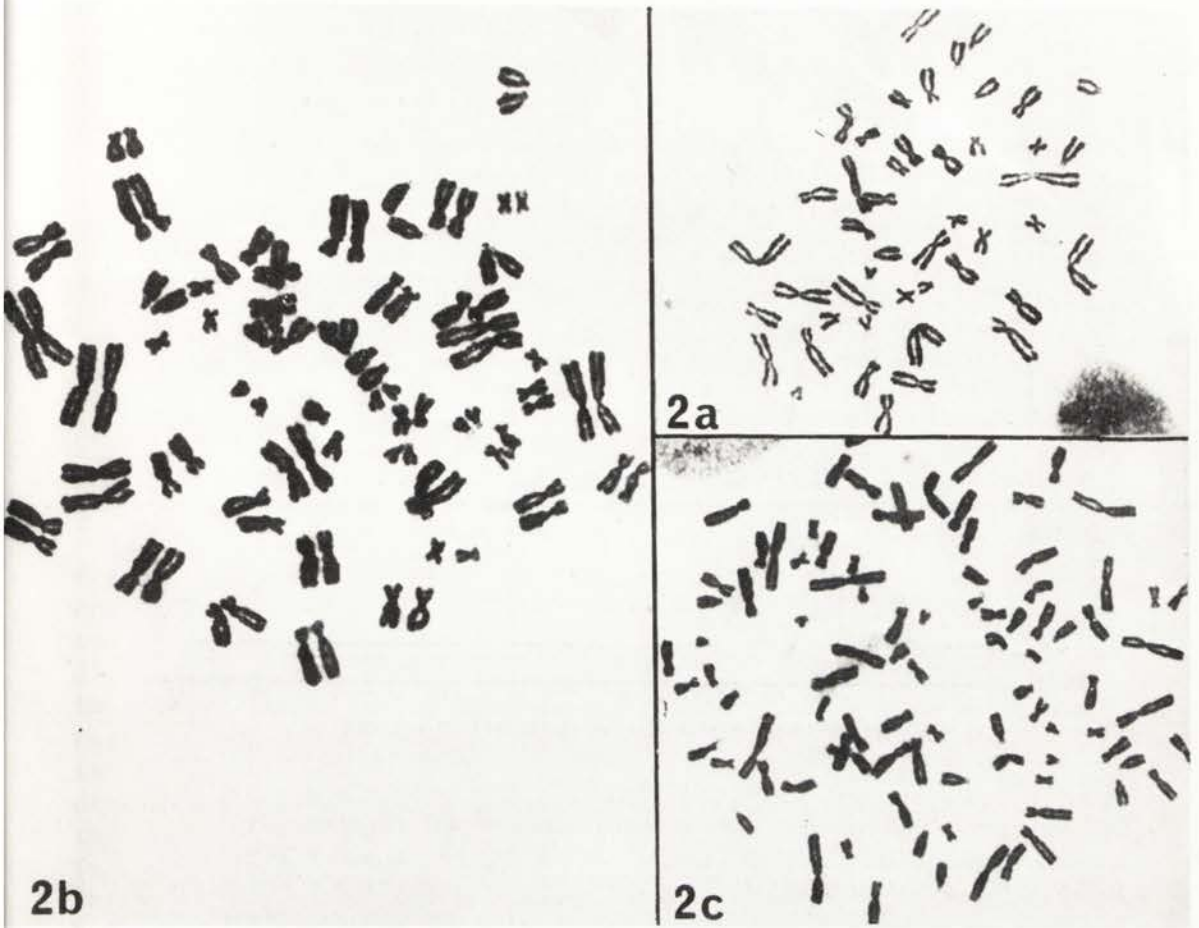


Figure 2. Three cell types found in cultures previously exposed to raised pH: (a) Normal 46 chromosome diploid metaphase. 3,000 x. (b) 'Endoreduplicated' cell where each chromosome in (a) is represented twice - such pairs of chromosomes are described as diplochromosomes. 4,000 x. (c) Polyploid cell with 92 chromosomes (4N). 3,000 x.

of the metaphases measured, the 'normal' metaphases and 'endoreduplicated' metaphases (on the same slide) had the same DNA content (Figure 4).

In order to demonstrate that the diplochromosomes had at least one DNA double helix per chromatid, cultures were labelled with tritiated thymidine one hour before exposure to raised pH and the metaphases at four hours were examined by autoradiography. All four chromatids of a pair of diplochromosomes were labelled, thus each chromatid contained at least one DNA helix.

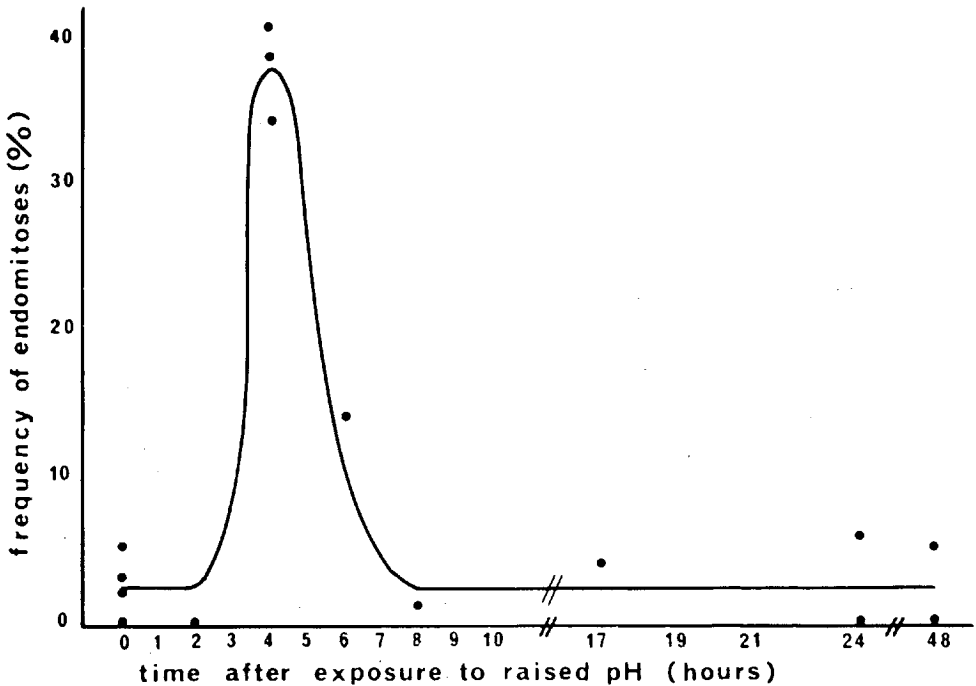


Figure 3. Frequency of 'endoreduplicated' cells after replacement of cultures in 5% CO₂. Values for control cultures (i.e. those maintained in 5% CO₂) are given as 0 hours. Each point represents a single experiment and all experiments were done on separate cultures of amniotic fluid.

Polyploids

Polyploid metaphases were observed in high frequency in cultures harvested at 17½ hours and later (Figure 5). Since the peak frequency of 38.46% was observed 24 hours after exposure to raised pH, and the cell cycle time is approximately 20 hours, it seems likely that the polyploid nuclei are related to the 'endoreduplicated' metaphases of the previous cell cycle. In Figure 6 the eight theoretical cell derivatives of an 'endoreduplicated' metaphase are listed - for convenience, split chromosomes are described here as unineme, and unsplit (or restored) as bineme; it is not implied that the evidence presented can distinguish between binemy and higher degrees of repetition.

Four types of metaphases are observed in cultures 24 hours after their exposure to raised pH: 41% are 4C, diploid (normal metaphases); 6% are 4C diplochromosomal; 38.5% are 8C tetraploid and 13% are tetraploid and diplochromosomal. Since the

TABLE 2

Frequency of labelled metaphases in cultures labelled with 1 μ Ci 3 H-TdR 10 min before exposure to raised pH

No. of cells	Time after return to 5% CO ₂	% metaphase labelled	% endomitosis labelled	% synthesis + G ₂ labelled
66	2 hours	0	0	20.2
248	4 hours	5.9	0	31.6
274	6 hours	89.6	27.0 8.1 (late)	31.5

frequency of the first type is less than in the control cultures and the frequency of the second is not greater than in the controls, it is unlikely that either of these represent daughter cells of diplochromosomal cells. However, since the other two types occur in control cultures with mean frequencies of only 4% and 0.2% respectively, and at 24 hours together have a frequency of 52%, it is most likely that they are the daughter cells. Both these types of metaphases can be formed only if 'restitution' occurs, but the two types may be formed by different patterns of DNA synthesis. Tetraploid cells are only formed by a DNA synthesis (Figure 6a) which leads to the restoration of the bineme chromosome structure but tetraploid, diplochromosomal cells may be formed by either type of synthesis. It is preferable to suggest that both types are formed by synthesis (a) and that the diplochromosomal cells are formed by splitting of the newly reformed bineme chromosomes. But, if this is so, the newly restored bineme chromosomes show a temporary instability which is not apparent at 48 hours.

DISCUSSION

Cells derived from amniotic fluid are difficult to grow; they are sensitive to antibiotics and fungicides, to changes in medium and above all to changes in pH. It is the last feature which has made these cells a valuable tool, since the inability of the cells to buffer against changes in external pH allows the manipulation of the internal cellular milieu.

Exposure of actively growing cultures to raised pH (not greater than pH 7.6) induces 'endoreduplication' and polyploidy in subsequent metaphases. However, close examination of these

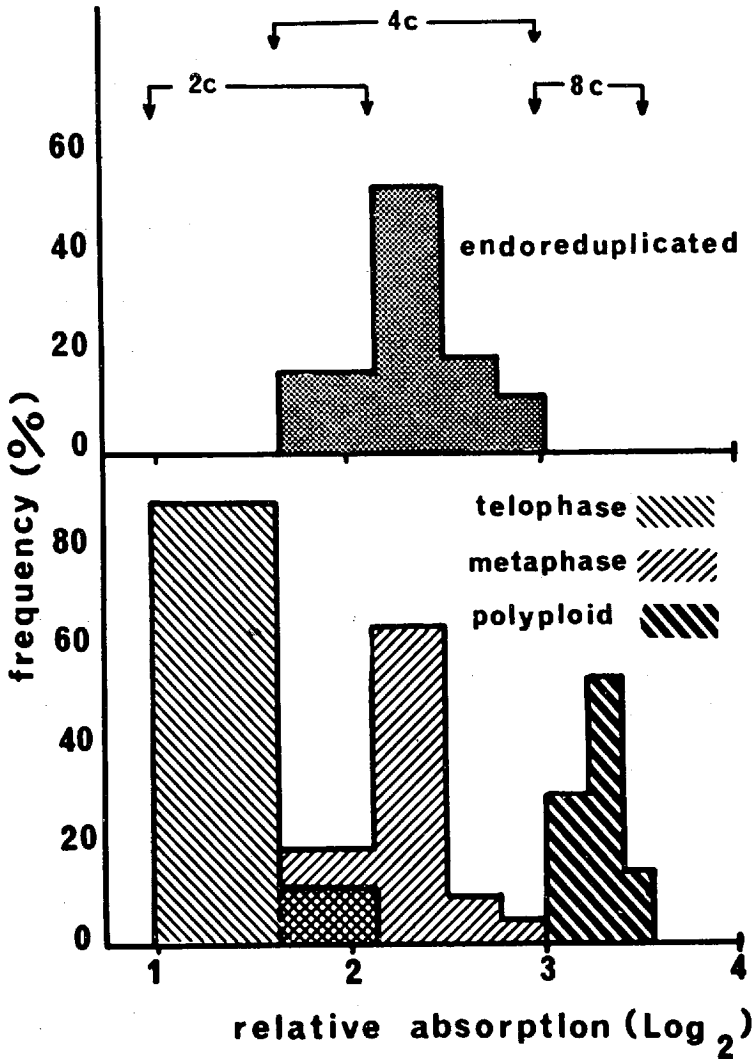


Figure 4. Relative absorption of Feulgen stained slides at 5600 Å. Results are expressed in log₂ so that each doubling of absorption is a unit increase. The 'endoreduplicated' metaphases have the same absorption as the diploid metaphases.

metaphases has indicated that they are different from colchicine induced endomitoses. Three sources of data have shown that apparently 'endoreduplicated' cells are derived by splitting the chromosomes of diploid metaphases: (i) the time of induction - these studies suggested that cells were at early G₂ at the time of exposure to raised pH; (ii) autoradiographic studies confirmed the time of induction and (iii) measurement of DNA content showed

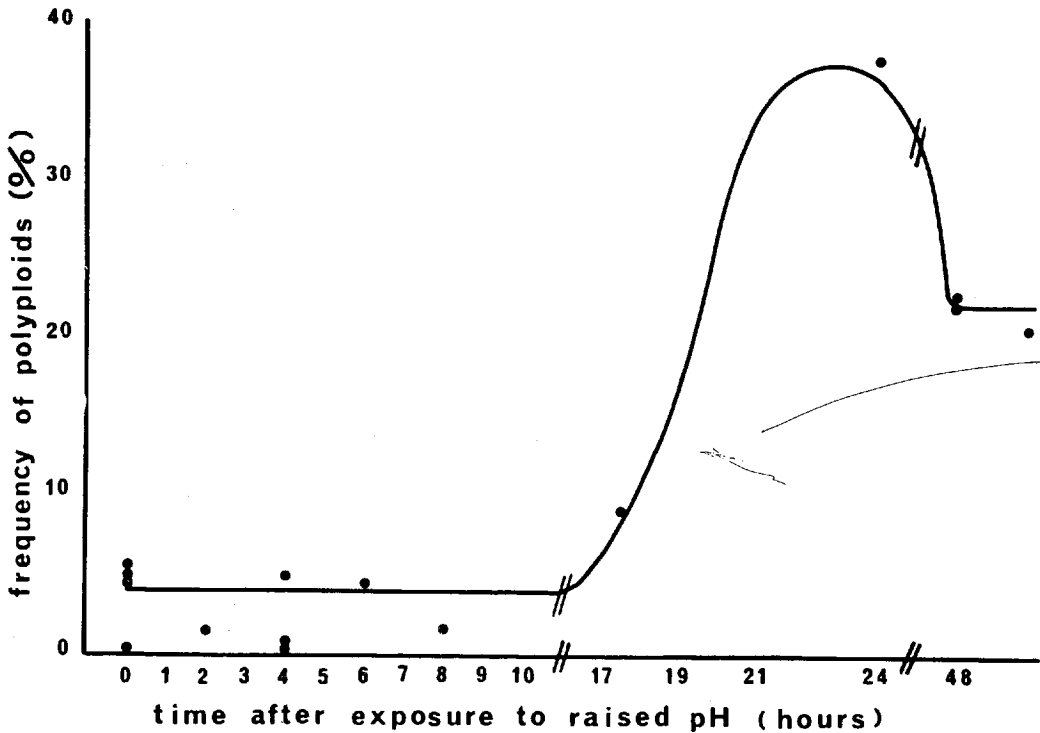


Figure 5. Frequency of polyploid cells after replacement of cultures in 5% CO₂. Values for control cultures are given as 0 hours.

that 'endoreduplicated' metaphases have the same DNA content as the diploid metaphases. Thus, there is no doubt that the diplochromosomes observed 4-6 hours after raised pH are each half-chromosomes and further, autoradiographic studies have shown that the chromatids of each half-chromosome contain at least one DNA double helix.

Studies of the next cell generation following splitting indicate that metaphase cells with split chromosomes do not enter anaphase. Failure to enter anaphase may reflect some insufficiency of the split chromosomes to attach to a spindle or may merely indicate that the rise in pH has inhibited the formation or function of the mitotic apparatus. Whichever the cause, analysis of cell frequencies at 24 hours indicates that following splitting, restitution nuclei are formed and DNA synthesis proceeds in such a way as to restore the bineme structure of chromatids in the now polyploid cells.

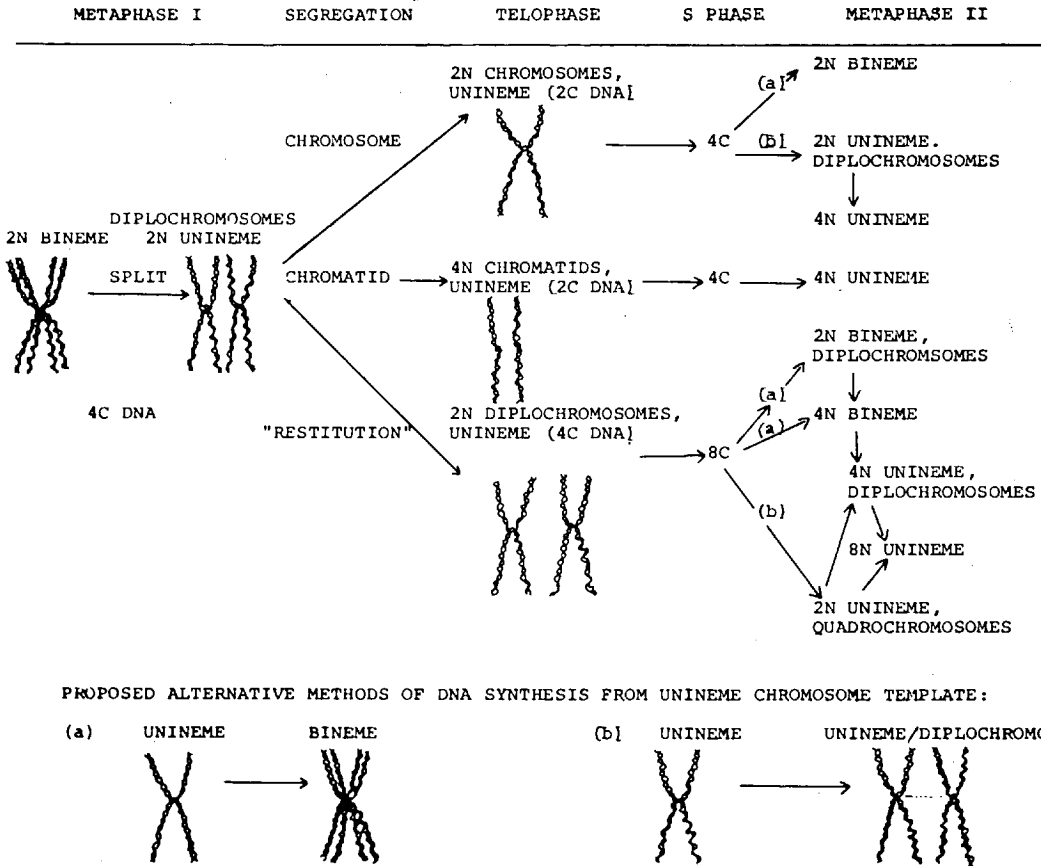


Figure 6. Theoretical, alternative behaviour of 'endoreduplicated' metaphases and the expected cell types following one round of DNA replication

Previous studies of chromosome structure have been unable to provide irrefutable evidence for either a unineme or polyneme chromosome, and it has been said (Prescott, 1970) that 'proof of a bipartite structure for the unduplicated chromosome is not likely to occur until it has been shown that the units can be separated to the degree that chromatids of the metaphase chromosome are normally separable'. In this study, evidence has been presented which demonstrates that in cells derived from amniotic fluid a rise in the pH of the medium can induce chromosome splitting; and further that each split chromatid contains at least one DNA helix. Thus, unsplit chromosomes must be at minimum bineme structures.

REFERENCES

- Bajer, A. 1965. Subchromatid structure of chromosomes in the living state. *Chromosoma* 17: 291.
- Gall, J.G. 1963. Kinetics of deoxyribonuclease action on chromosomes. *Nature* 198: 36.
- Kaufmann, B.P., H. Gay and M.R. McDonald. 1960. Organizational patterns within chromosomes. *Int. Rev. Cytol.* 9: 77.
- Prescott, D.M. 1970. The structure and replication of eukaryotic chromosomes. *Adv. Cell Biol.* 1: 57.
- Ris, H. 1966. Fine structure of chromosomes. *Proc. Roy. Soc. London B* 164: 245.
- Taylor, J.H. 1963. The replication and organization of DNA in chromosomes. In *Molecular Genetics* (J.H. Taylor, ed.) p. 65. Academic Press, New York.
- Taylor, J.H., P.S. Woods and W.L. Hughes. 1957. The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labeled thymidine. *Proc. Nat. Acad. Sci. USA* 43: 122.
- Trosko, J.E. and S. Wolff. 1965. Strandedness of *Vicia faba* chromosomes as revealed by enzyme digestion studies. *J. Cell. Biol.* 26: 125.

Ribosomal gene amplification in oocytes and in primitive eukaryotes

Joseph G. Gall

Extrachromosomal copies of the genes coding for ribosomal RNA exist not only in the oocytes of many higher organisms, but also in the vegetative nuclei of the protozoan, *Tetrahymena pyriformis* and probably some other primitive eukaryotes. In *Tetrahymena*, as in oocytes, the extrachromosomal genes represent a true amplification relative to the number of genes present in the basic genome. The replication of the amplified rDNA of oocytes involves a rolling circle mechanism, as shown by electron microscopy and autoradiography. The amplification process in *Tetrahymena* has not yet been studied. The integration of multiple rDNA copies into the chromosomes of higher eukaryotes may be an evolutionarily derived phenomenon. The integration may be somewhat unstable, so that extrachromosomal copies are regularly available for various types of amplification phenomena.

rDNA AMPLIFICATION IN OOCYTES

Ribosomal gene amplification involves a specific extrachromosomal replication of the genes coding for ribosomal RNA. It has been most extensively studied in the oocytes of the South African clawed toad, *Xenopus laevis* (Brown and Dawid, 1968; Gall, 1968). The major period of synthesis occurs during the pachytene stage of meiosis in oocytes, and results in the production of 25-30 pg of extra rDNA, nearly ten times the mass of the haploid genome (Perkowska et al., 1968). Several reviews on amplification have appeared in the past few years (e.g. Gall, 1969; Hourcade et al., 1974; Tobler, 1974). We shall be concerned here primarily with two questions: the origin of the amplified rDNA, and the nature of its replication.

ORIGIN OF AMPLIFIED rDNA

Any discussion of the origin of the amplified rDNA must begin with a consideration of its inheritance. Specifically, it must be known whether the rDNA is carried extrachromosomally from generation to generation through the egg (episome model), or whether it is produced anew each generation using the chromosomal rDNA as template (chromosome copy model). A definitive answer to this question was provided by the elegant experiments of Brown and Blackler (1972),

utilizing species hybrids between *X. laevis* and *X. mulleri*. The rDNA of these two species can be distinguished easily on the basis of RNA-DNA hybridization experiments. Although the regions coding for the rRNA are similar or identical in the two species, the 'spacer' regions between the transcribed genes are quite different. Complementary RNA synthesized from the spacer region of one species will hybridize very little if at all with the spacer regions of the other species. Using the complementary RNAs as probes, Brown and Blackler showed that the amplified rDNA in the oocytes of F1 hybrid female toads was invariably *laevis*, regardless of which parent was *laevis* and which was *mulleri*. The exclusive amplification of *laevis* rDNA was unexpected, since the F1 animals carried both *laevis* and *mulleri* rDNA in their somatic nuclei. But the cross in which the male parent was *laevis* ruled out strictly maternal inheritance of the amplified rDNA, as would have occurred if the rDNA were passed as an extrachromosomal episome in the egg.

Knowing that amplification proceeds by a chromosome copy mechanism, we can examine more closely the origin of the extrachromosomal rDNA. The question we wish to answer is, 'When and how does the first extrachromosomal rDNA copy arise?' Some limited statements can be made about 'when' from cytological studies, but virtually nothing is known about 'how'. Indeed the available information suggests that higher eukaryotes provide particularly intractable systems in which to study the latter question, at least by biochemical means.

The first *in situ* nucleic acid hybridization studies showed that amplification begins well before the oocyte stage in amphibians, although by far the bulk of extrachromosomal synthesis occurs after the start of the first meiotic prophase (Gall and Pardue, 1969). Amplified rDNA can be recognized in oogonia of *Xenopus* by comparing the amount of *in situ* hybridization in oogonial nuclei with that in adjacent follicle cell nuclei. Although accurate estimates are not possible because of differences in nuclear size and shape, and the general unreliability of *in situ* hybrids for quantification, it is probable that oogonia contain over 10 times as much rDNA as normal diploid nuclei. By comparison a late pachytene oocyte contains approximately 2,000 times the diploid rDNA content. During oogonial mitoses the extra DNA appears as clumps of staining material among the dividing chromosomes. It is probable that the amplified rDNA is more or less evenly divided between the products of the oogonial mitoses, since amplified rDNA is detectable during interphase in all the oogonia of a cluster. Recently Kalt and Gall (1974) studied stages as early as the primordial germ cells by *in situ* hybridization. The first, nondividing germ cells that populate the genital ridge hybridize with rRNA to about the same extent as adjacent somatic nuclei (they average 1.3 times as many silver grains in autoradiographs,

although this may be due to their more flattened shape). At the time that the primordial germ cells begin dividing, and hence become oogonia by definition, they begin to show extra rDNA by *in situ* hybridization. If the first extrachromosomal copies arise at this time, they will be very difficult to study by biochemical means. Unfortunately there are only a few dozen primordial germ cells per embryo and there is no method by which they can be separated in large numbers from somatic cells.

The existence of amplified rDNA before the oocyte stage is more dramatically demonstrated in other organisms. For instance, in his classic study of the beetle, *Dytiscus marginalis*, Giardina (1901) showed that extra 'chromatin' is prominent throughout the oogonial divisions. At each mitosis this chromatin (=rDNA) is aggregated into a horseshoe or doughnut shaped mass that moves to one pole only. After four oogonial divisions there are sixteen cells in a cluster, only one of which carries the large mass of rDNA. The one cell with the amplified rDNA becomes an oocyte, while the other fifteen become nurse cells. As in *Xenopus*, the cytological data in *Dytiscus* demonstrate that amplification begins no later than the earliest oogonial stages.

An interesting sidelight to the amplification story is the fact that spermatogonia of *Xenopus* are amplified to about the same extent as oogonia, as shown by *in situ* hybridization (Pardue and Gall, 1972; Kalt and Gall, 1974). The amplified rDNA is apparently lost at the time that spermatogonia become spermatocytes, since no unusual hybridization is detectable after the beginning of meiosis.

Essentially nothing is known about the origin of the first extrachromosomal rDNA copies. By analogy with bacterial and viral systems one can easily postulate mechanisms such as recombination, enzymatic excision, or special replications which would allow one or more rDNA units to become free of the chromosomes. None of these models is readily testable because of the scarcity of the appropriate cells. A mechanism that has received much discussion is that proposed by Crippa and Tocchini-Valentini (1971). They suggested that the amplified rDNA is copied by a reverse transcriptase enzyme from special rRNA transcripts containing the usually nontranscribed spacer regions. This model has a certain appeal since it would account for the first rDNA copies as well as the major pachytene synthesis. However, it should be noted that all the experimental data *pro* and *con* the reverse transcriptase theory refer to rDNA synthesis in the *oocyte* stage (Bird et al., 1973b; Tocchini-Valentini et al., 1974). We have essentially no biochemical evidence concerning the oogonial and primordial germ cell stages.

REPLICATION OF AMPLIFIED rDNA

The possibility that amplified rDNA exists as circular molecules was suggested as early as 1965 by the work of Miller (1966) and Kezer (1965). These investigators showed that amphibian oocyte nucleoli may take on the form of beaded necklaces easily visible in the light microscope, and that these necklaces were extensively fragmented when treated with DNase. Later Miller and Beatty (1969) examined the same material by electron microscopy after partial dissolution in very dilute saline solutions. The rDNA with attached precursor rRNA transcripts was beautifully demonstrated. Although the chains of adjacent transcribing regions were usually tangled and could not be followed for their entire length, the absence of free ends was taken to imply circularity of the rDNA.

Recently Hourcade et al. (1973;1974) made the important discovery that purified rDNA extracted from amplifying stages of *Xenopus* contained circular molecules observable by electron microscopy. The molecules found by Hourcade et al. were much smaller than those studied by Miller and Kezer. The smallest were only 7.5×10^6 daltons in mass, equivalent to one coding region for precursor rRNA plus the accompanying spacer. Larger circles were integral multiples of this monomer size containing 2, 3, 4 or more coding regions. Some very large circles were found containing up to 15 or more coding units. Since circles of this size are very likely to be broken by shear during the preparative steps, we do not really know what fraction of the total rDNA exists as circles, nor what is the relative frequency of the different sizes. As isolated and spread for electron microscopy most of the rDNA (about 98%) exists as linear molecules, but these *could* all be broken circles.

Gall and Rochaix (1974) have recently studied the amplified rDNA from oocytes of two beetles, *Colymbetes fuscus* and *Dytiscus marginalis*. As in *Xenopus*, circular molecules were present, and these fell into size classes that were integral multiples of a basic unit. The unit size was 10.0×10^6 daltons in *Colymbetes* and 19.2×10^6 daltons in *Dytiscus*. In one sample of rDNA 19% of the molecules were circular, suggesting that an even higher percentage may be so *in vivo*.

Hourcade et al. (1973;1974) observed not only circles but also lariats (circles with tails) in the rDNA of *Xenopus*, and they suggested that such lariats might be 'rolling circle' replication forms. In the rolling circle model of DNA replication a single replication fork moves continuously around the circle, spinning out an ever lengthening linear tail. On this model one should expect to find tails longer than the circles to which they are attached. These were seen by Hourcade et al. and also later by Bird et al. (1973a) and Rochaix et al. (1974). It is possible for a broken 'Cairn's form' molecule

to appear as a circle with a tail, but in these cases the tail will always be shorter than the circle. Only two unbroken Cairn's forms were seen by Hourcade et al. and none by Bird et al. The rolling circle hypothesis has recently been tested by Bird et al. using electron microscope autoradiography. By labeling the amplifying rDNA with ^3H -thymidine and then preparing E.M. autoradiographs of the spread rDNA molecules, they were able to pick out the replicating molecules on the basis of their associated silver grains. As the pulse time was decreased the frequency of labeled molecules which were rolling circles went up to a maximum of 18%. The rolling circle model of replication was further tested by examining the distribution of label on the molecules. In very short pulse experiments label should be preferentially located in the region where the tail joins the circle. In longer pulses it should occur over the circle and in the proximal portion of the tail. And in pulse-chase experiments the label should be located only near the free end of the tail. These predictions were borne out by appropriate experiments in which amplifying oocytes were labeled *in vitro* with ^3H -thymidine.

rDNA AMPLIFICATION IN *TETRAHYMENA*

During vegetative growth the macronucleus of *Tetrahymena* contains 500-1000 peripheral-nucleoli reminiscent of the multiple oocyte nucleoli of an amphibian, although on a much reduced scale. Charret (1969) showed that these nucleoli contain cytochemically demonstrable DNA and incorporate ^3H -thymidine independently of the rest of the macronucleus. Charret's autoradiographic study was complemented by the biochemical work of Engberg et al. (1972), who demonstrated an independent replication of the rDNA during a nutritional shift-up. These facts originally prompted our search for extrachromosomal rDNA in this organism. Unknown to us at the start of our work, Engberg et al. (1974) had already discovered that the nucleolar DNA was of low molecular weight, even when isolated under conditions of minimal shear. Their observations provided the first strong evidence for the extrachromosomal nature of the rDNA in *Tetrahymena*.

The state of the macronuclear rDNA has been studied by electron microscopy (Gall, 1974). A purified rDNA fraction was prepared from *Tetrahymena pyriformis*, amiconucleate strain GL, by three cycles of ultracentrifugation in CsCl gradients. The rDNA is more dense than the bulk DNA (Engberg et al., 1972), the densities being 1.699 g/cm^3 and 1.689 g/cm^3 respectively. As shown by electron microscopy of material spread by the Kleinschmidt technique (Davis et al., 1971), nearly all of the rDNA consisted of molecules 12.6×10^6 daltons in mass. Of these approximately 90% were linear

and most of the rest were circular (some being supercoiled circles). There were a few dimer and multimer linear and circular molecules, and a few lariats, most of which had monomer length tails attached to a monomer length circle. Denaturation mapping showed that all of the linear molecules were similar or identical in sequence, and therefore had not arisen by random breakage of circles. For these reasons it is believed that the rDNA *in vivo* probably contains both linear and circular molecules.

An experiment in which purified rDNA was hybridized with rRNA showed an RNA/DNA ratio of 12.6% at saturation. This value is only slightly lower than that predicted on the assumption that each monomer rDNA molecule codes for one 17S and one 25S rRNA molecule (the combined molecular weight of the two rRNAs is 2.0×10^6 (Loening, 1968) giving an expected saturation value of $2.0 \times 10^6 / 12.6 \times 10^6 = 15.9\%$).

Thus the rDNA in the macronucleus during vegetative growth consists essentially of a homogeneous population of linear and circular molecules. Approximately 1/3 of the molecule codes for 17S and 25S rRNA and the remainder is analogous to the 'spacer' region described for the rDNA of *Xenopus* and other eukaryotes (Birnstiel et al., 1968; Wensink and Brown, 1971). Of the remaining 2/3 of the molecule, part is transcribed, since a high molecular weight precursor rRNA molecule has been demonstrated (Leick, 1969). It is not known whether rRNA is transcribed from both the linear and circular molecules. This question could be investigated by preparing material according to the technique of Miller and Beatty (1969).

The number of rDNA copies in the macronucleus can be calculated from published data on the macronuclear DNA content and the percent of macronuclear DNA homologous to rRNA. In strain GL Engberg and Pearlman (1972) reported that rRNA was homologous to 0.3% of the DNA. The macronuclear DNA content has been variously estimated between 10 and 20 pg. For the present calculation I shall use the value of 10.16 pg for the G₁ stage (Woodard et al., 1972), although this was determined on syngen I and may not be exactly the same as in strain GL. These data predict about 10,000 rDNA copies per macronucleus. If these are uniformly distributed among the 500-1,000 peripheral nucleoli, then there are approximately 10 to 20 rDNA molecules per nucleolus.

The haploid DNA content of *Tetrahymena* has been determined to be 0.22 pg by microdensitometric measurements on G₂ (4C) micronuclei (Woodard et al., 1972). Recently Yao and Gorovsky (personal communication) have shown that the kinetic complexity of macro- and micronuclear DNA is very nearly the same, corresponding to a

unit genome of about 0.2 pg. The unit genome, defined by reassociation kinetics, and the haploid DNA content measured by microdensitometry, are, therefore, essentially the same. To a first approximation we may regard the G_1 micronucleus as containing 2 genomes and the G_1 macronucleus 50 genomes. Since there are 50 genomes in the macronucleus and 10,000 rDNA molecules, there are approximately 200 rDNA copies per haploid genome.

The fraction of the genome devoted to rDNA, as well as the genome size itself, are similar to values found in other eukaryotes. For instance, *Drosophila melanogaster* has a genome size of 0.18 pg and its rDNA constitutes 0.3 to 0.4%. Thus, although the ribosomal cistrons are extrachromosomal in the *Tetrahymena* macronucleus, there was no *a priori* reason to believe that the rDNA was amplified relative to the basic genome. Nevertheless, that appears to be the case, as shown recently by Yao et al. (1974). These workers have purified micro- and macronuclear DNA (syngen I) and have determined the amount of rRNA bound by each in saturation hybridization experiments. Using 25S rRNA they determined a saturation value of 0.2% for macronuclear DNA, but only 0.02% for micronuclear. With a mixture of 25S and 17S rRNA the values were 0.3% and 0.03%. The macronuclear value is in good agreement with the earlier estimate of Engberg and Pearlman (1972) for strain GL, and corresponds to about 200 copies per genome in the macronucleus. On the other hand, the micronuclear value corresponds to no more than 20 rDNA copies per genome. Since micronuclear preparations invariably contain some macronuclei, and each macronucleus contains on the average 20x as much DNA as a micronucleus, the micronuclear DNA is never pure. In the experiments of Yao et al. the contamination was estimated as 5-15% by mass. This means that much of the hybridization of rRNA to the micronuclear DNA preparation could be accounted for by the contaminating macronuclear DNA. Thus the number of rDNA copies per haploid genome in the micronucleus is certainly less than 20, and could be as low as one or two.

From the hybridization experiments it is, of course, impossible to tell whether the micronuclear rDNA copies are integrated into the chromosomes, or free, as in the macronucleus. Whatever the state of the micronuclear rDNA, we can say that the rDNA in the macronucleus is amplified relative to that in the micronucleus. Because of the history of the micronucleus during conjugation we can safely predict that amplification occurs shortly after the double fertilization characteristic of *Tetrahymena* and other ciliates. At the time of conjugation the micronucleus goes through meiosis to produce gamete nuclei. After fertilization the new diploid micronucleus gives rise not only to the micronucleus of

the next generation, but also, through polyploidization, to the new macronucleus. In the meantime the old macronucleus disintegrates. It is during the period of macronuclear formation that amplification of the rDNA probably takes place. The molecular aspects of amplification, such as the possible existence of rolling circle replication, should be easy to study, since conjugation can be induced in mass populations of *Tetrahymena* by appropriate manipulation of the culture conditions (Bruns and Brussard, 1974).

Replication of the rDNA during vegetative fission has been shown to be semi-conservative by experiments involving BUdR incorporation (Engberg et al., 1972). Since this replication presumably involves a simple doubling of the rDNA, replication might take one of several forms. Clearly, a fork might arise at one end of a linear molecule and proceed to the other, or arise internally and move bidirectionally (Wolfson et al., 1972). No such molecules have yet been seen by electron microscopy. If the circular and lariat molecules are involved in a rolling circle mechanism, then we should predict that each circle would 'roll out' only a single monomer product. Those lariat forms which have been seen often had tails the same length as the circle. However, such evidence is unconvincing until coupled with labeling experiments. It has been demonstrated by Engberg et al. (1972) that rDNA and mitochondrial DNA replication are preferentially initiated when cultures are re-fed after a 24-hour period of starvation. Their technique should prove especially useful in studying the molecular aspects of rDNA replication in vegetative cells.

rDNA AMPLIFICATION IN OTHER ORGANISMS

The question naturally arises whether the situation in *Tetrahymena* is unique or represents a more general condition, say, among primitive eukaryotes. There are strong suggestions that other organisms possess extrachromosomal rDNA. The most compelling evidence comes from biochemical studies on the ciliate, *Stylonychia*, and the slime mold, *Physarum*.

In *Stylonychia* Prescott et al. (1973) have shown that the macronuclear DNA consists entirely of short molecules, 0.2 to 2.2 μm in length by electron microscopy. On a sucrose gradient these molecules sediment as a main peak with a sedimentation coefficient of about 10S and a minor peak of about 14S. Prescott and Murti (1974) showed by RNA-DNA hybridization experiments that the rDNA is located preferentially in the 14S peak, as would be expected from the known size of the rRNA precursor molecule. The existence of extrachromosomal rDNA in *Stylonychia* is, therefore, well established. However, the fact that all the macronuclear DNA consists of 'gene-sized' pieces complicates the issue. Whether the rDNA is

amplified as well as being extrachromosomal could be examined in *Stylonychia*, as in *Tetrahymena*, by carrying out RNA-DNA hybridization experiments with micronuclear DNA. Electron microscopic observations on micronuclear DNA show very long molecules and no evidence that the bulk of the DNA is of low molecular weight.

In the slime mold, *Physarum*, Sonenschein and Holt (1968), Newlon et al. (1973) and Zellweger et al. (1972) have provided two kinds of evidence that the rDNA is of low molecular weight, and therefore possibly extrachromosomal. If *Physarum* cells are lysed with sodium lauryl sarcosinate on top of a sucrose gradient and then centrifuged, the bulk of the nuclear DNA sediments faster than the rDNA. Also the rDNA can be partially purified in a single-step salt precipitation using the method of Hirt (1967), which was developed to separate low molecular weight viral DNA from the nuclear DNA of virus infected cells. Assuming that the *Physarum* rDNA is mainly extrachromosomal, one would like to know whether chromosomal copies also exist. Unlike *Tetrahymena*, *Physarum* does not have micronuclei in which this question could be approached experimentally. It would be useful to ask, however, whether there are any nuclei in the life cycle that have a reduced rDNA content and from which vegetative nuclei arise by amplification. Attention should perhaps be focused on the transition from the spore to amoeba stage.

In the alga, *Acetabularia*, Trendelenburg et al. (1974) have presented data consistent with the existence of amplified, extrachromosomal rDNA, but which are not conclusive. They have shown that the prominent nucleoli contained in the giant vegetative nucleus can be studied successfully by the Miller spreading technique. The nucleoli contain long stretches of tandemly arranged units very similar to the transcribing rDNA units of *Xenopus*. It is clear that the rDNA units do not exist as separate low molecular weight entities, as they do in the cases so far discussed, but they may nevertheless be both amplified and extrachromosomal. An extrachromosomal nature is implied, though not proven, by the nucleolar location. Amplification is suggested by the enormous number of such units per nucleus, estimated by Trendelenburg et al. as about 13,000. Once again, amplification can only be defined relative to some other nuclear state. In this case, it would be interesting to examine the rDNA in the spores or in the gametes released from them.

Other simple eukaryotes may well be worth investigating. It has been reported that rDNA replication in *Chlamydomonas* is unusual, involving an overproduction and then breakdown during each vegetative mitosis (Howell, 1972).

Despite the paucity of information it might be useful to set up a model of rDNA evolution as a guide for further experimentation. We can imagine that the earliest eukaryotes possessed only a few rDNA cistrons, integrated into chromosomes as in the case of present day bacteria. Because of the demands for rRNA imposed by their large cytoplasmic volume, these primitive eukaryotes required more rDNA templates. These templates were produced as multiple extrachromosomal copies, perhaps replicated during vegetative mitosis but discarded and then regenerated at the time of meiosis and fertilization. *Tetrahymena* may illustrate this condition in a diagrammatic fashion, the micronucleus containing only integrated copies, the macronucleus containing largely amplified, extrachromosomal copies, perhaps along with a small fraction of chromosomally integrated genes in the polyploidized chromosomes. The areas for future investigation are reasonably clear. We need to know whether the micronuclear copies are integrated or not. We need to know whether amplification occurs during the change from micronucleus to macronucleus. In other organisms we need to know whether extrachromosomal rDNA exists during vegetative mitosis. If extrachromosomal rDNA is found, we should like to know if it is amplified in amount relative to some other stage, such as the germinal nucleus. Finally a survey of primitive eukaryotes, including even multicellular forms, would be interesting to see at what level of complexity the switch from multiple extrachromosomal to multiple chromosomally integrated rDNA cistrons was made.

As a final flight of fancy, we should perhaps examine more closely the whole question of rDNA integration in higher eukaryotes. Certainly a number of puzzling facts are known which suggest that rDNA replication is not strictly 'normal' even in somatic cells. There is the case of under-replication in polytene cells of *Drosophila* (Tartof, 1971; Hennig and Meer, 1971; Spear and Gall, 1973), as well as the curious phenomenon of 'magnification' whereby stocks of flies deficient in ribosomal cistrons regain them over a period of only a few generations (Ritossa, 1968). Ritossa (1972) has argued that magnification could involve an extrachromosomal rDNA state, although biochemical data are not yet available in this case. Perhaps a certain number of rDNA copies are regularly excised from chromosomes, and are in a sense always available for magnification or amplification phenomena, depending upon the peculiarities of the life cycle and the demand for ribosomal cistrons.

Finally, one might ask if the unusual chromosomal location of the rDNA is related to its state of integration. Lima-de-Faria (1973) has tabulated the position of the nucleolus or presumed nucleolar constrictions in 506 different species, and in the

overwhelming majority of these (86.6%) the constriction occurs near the end of the short arm of a chromosome. In cases where this location is not found, the nucleolus is often in a major heterochromatic region, as in the *Drosophila* X and Y chromosomes. Perhaps these positions confer some special ability on the rDNA to become excised or reintegrated into the linear continuity of the chromosome.

In summary, extrachromosomal copies of the genes coding for rRNA exist not only in the oocytes of some higher organisms, but in the vegetative nuclei of *Tetrahymena* and probably in other primitive eukaryotes. In *Tetrahymena* these extrachromosomal genes represent a true amplification relative to the number of genes present in the basic genome - in this case the micronucleus. Integration of multiple rDNA copies into the chromosomes of higher eukaryotes may be an evolutionarily derived phenomenon. The integration may be somewhat unstable, so that extrachromosomal copies are regularly available for various types of amplification phenomena.

ACKNOWLEDGMENTS

Support by Research Grant GM 12427 from the National Institute of General Medical Sciences. The technical assistance of Mrs C. Barney and Mrs S. Hodgson is gratefully acknowledged.

REFERENCES

- Bird, A., J.-D. Rochaix and A. Bakken. 1973a. The mechanism of gene amplification of *Xenopus laevis* oocytes. In *Molecular Cytogenetics* (B. Hamkalo and J. Papaconstantinou, ed.) p.49. Plenum Publ. Corp., N.Y.
- Bird, A., E. Rogers and M. Birnstiel. 1973b. Is gene amplification RNA-directed? *Nature New Biol.* 242: 226.
- Birnstiel, M., J. Speirs, I. Purdom, K. Jones and U.E. Loening. 1968. Properties and composition of the isolated ribosomal DNA satellite of *Xenopus laevis*. *Nature* 219: 454.
- Brown, D.D. and A.W. Blackler. 1972. Gene amplification proceeds by a chromosome copy mechanism. *J. Mol. Biol.* 63: 75.
- Brown, D.D. and I.B. Dawid. 1968. Specific gene amplification in oocytes. *Science* 160: 272.
- Bruns, P. and T. Brussard. 1974. Pair formation in *Tetrahymena pyriformis*, an inducible developmental system. *J. Exp. Zool.* In press.
- Charret, R. 1969. L'ADN nucléolaire chez *Tetrahymena pyriformis*: chronologie de sa répllication. *Exp. Cell Res.* 54: 353.
- Crippa, M. and G.P. Tocchini-Valentini. 1971. Synthesis of amplified DNA that codes for ribosomal RNA. *Proc. Nat. Acad. Sci. USA* 68: 2769.

- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *In Methods in enzymology* (L. Grossman and K. Moldave, ed.) Vol. 21 p.413. Academic Press, N.Y.
- Engberg, J. and R.E. Pearlman. 1972. The amount of ribosomal RNA genes in *Tetrahymena pyriformis* in different physiological states. *Eur. J. Biochem.* 26: 393.
- Engberg, J., D. Mowat and R.E. Pearlman. 1972. Preferential replication of the ribosomal RNA genes during a nutritional shift-up in *Tetrahymena pyriformis*. *Biochim. Biophys. Acta* 272: 312.
- Engberg, J., J.R. Nilsson, R.E. Pearlman and V. Leick. 1974. Induction of nucleolar and mitochondrial DNA replication in *Tetrahymena pyriformis*. *Proc. Nat. Acad. Sci. USA* 71: 894.
- Gall, J.G. 1968. Differential synthesis of the genes for ribosomal RNA during amphibian oögenesis. *Proc. Nat. Acad. Sci. USA* 60: 553.
- . 1969. The genes for ribosomal RNA during oögenesis. *Genetics* 61 (Suppl.): 121.
- . 1974. Free ribosomal RNA genes in the macronucleus of *Tetrahymena*. *Proc. Nat. Acad. Sci. USA*. In press.
- Gall, J.G. and M.L. Pardue. 1969. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Nat. Acad. Sci. USA* 63: 378.
- Gall, J.G. and J.-D. Rochaix. 1974. The amplified ribosomal DNA of Dytiscid beetles. *Proc. Nat. Acad. Sci. USA* 71: 1819.
- Giardina, A. 1901. Origine dell'oocite e della cellule nutrice nel *Dytiscus*. *Int. Mschr. Anat. Physiol.* 18: 418.
- Henning, W. and B. Meer. 1971. Reduced polyteny of ribosomal RNA cistrons in giant chromosomes of *Drosophila hydei*. *Nature New Biol.* 233: 70.
- Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* 26: 365.
- Hourcade, D., D. Dressler and J. Wolfson. 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. *Proc. Nat. Acad. Sci. USA* 70: 2926.
- , ———, ———. 1974. The nucleolus and the rolling circle. *Cold Spring Harbor Symp. Quant. Biol.* 38: 537.
- Howell, S.H. 1972. The differential synthesis and degradation of ribosomal DNA during the vegetative cell cycle in *Chlamydomonas reinhardi*. *Nature New Biol.* 240: 264.
- Kalt, M.R. and J.G. Gall. 1974. Observations on early germ cell development and premeiotic ribosomal DNA amplification in *Xenopus laevis*. *J. Cell Biol.* In press.

- Kezer, J., cited in W.J. Peacock. 1965. Chromosome replication *Nat. Cancer Inst. Mono.* 18: 101.
- Leick, V. 1969. Formation of subribosomal particles in the macronuclei of *Tetrahymena pyriformis*. *Eur. J. Biochem.* 8: 221.
- Lima-de-Faria, A. 1973. Equations defining the position of ribosomal cistrons in the eukaryotic chromosome. *Nature New Biol.* 241: 136.
- Loening, U.E. 1968. Molecular weights of ribosomal RNA in relation to evolution. *J. Mol. Biol.* 38: 355.
- Miller, O.L. 1966. Structure and composition of peripheral nucleoli of salamander oocytes. *Nat. Cancer Inst. Mono.* 23: 53.
- Miller, O.L. and B.R. Beatty. 1969. Visualization of nucleolar genes. *Science* 164: 955.
- Newlon, C.S., G.E. Sonenshein and C.E. Holt. 1973. Time of synthesis of genes for ribosomal ribonucleic acid in *Physarum*. *Biochemistry* 12: 2338.
- Pardue, M.L. and J.G. Gall. 1972. Molecular cytogenetics. In *Molecular genetics and developmental biology* (M. Sussman, ed.) p.65. Academic Press, N.Y.
- Perkowska, E., H.C. Macgregor and M.L. Birnstiel. 1968. Gene amplification in the oocyte nucleus of mutant and wild-type *Xenopus laevis*. *Nature* 217: 649.
- Prescott, D.M. and K.G. Murti. 1974. Chromosome structure in ciliated protozoans. *Cold Spring Harbor Symp. Quant. Biol.* 38: 609.
- Prescott, D.M., K.G. Murti and C.J. Bostock. 1973. Genetic apparatus of *Stylonychia* sp. *Nature* 242: 576 and 597.
- Ritossa, F.M. 1968. Unstable redundancy of genes for ribosomal RNA. *Proc. Nat. Acad. Sci. USA* 60: 509.
- . 1972. Procedure for magnification of lethal deletions of genes for ribosomal RNA. *Nature New Biol.* 240: 109.
- Rochaix, J.-D., A. Bird and A. Bakken. 1974. Ribosomal RNA gene amplification by rolling circles. *J. Mol. Biol.* In press.
- Sonenshein, G.E. and C.E. Holt. 1968. Molecular weight of mitochondrial DNA in *Physarum polycephalum*. *Biochem. Biophys. Res. Comm.* 33: 361.
- Spear, B.B. and J.G. Gall. 1973. Independent control of ribosomal gene replication in polytene chromosomes of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 70: 1359.
- Tartof, K.D. 1971. Increasing the multiplicity of ribosomal RNA genes in *Drosophila melanogaster*. *Science* 171: 294.
- Tobler, H. 1974. Occurrence and developmental significance of gene amplification. In *The Biochemistry of Animal Development* (R. Weber, ed.) Vol. 3. In press.

- Tocchini-Valentini, G.P., V. Mahdavi, R. Brown and M. Crippa. 1974. The synthesis of amplified ribosomal DNA. *Cold Spring Harbor Symp. Quant. Biol.* 38: 551.
- Trendelenburg, M., H. Spring, U. Scheer and W. Franke. 1974. Morphology of nucleolar cistrons in a plant cell, *Acetabularia mediterranea*. *Proc. Nat. Acad. Sci. USA*. In press.
- Wensink, P.C. and D.D. Brown. 1971. Denaturation map of the ribosomal DNA of *Xenopus laevis*. *J. Mol. Biol.* 60: 235.
- Wolfson, J., D. Dressler and M. Magazin. 1972. Bacteriophage T7 DNA replication: a linear replicating intermediate. *Proc. Nat. Acad. Sci. USA* 69: 499.
- Woodard, J., E. Kaneshiro and M.A. Gorovsky. 1972. Cytochemical studies on the problem of macronuclear subnuclei in *Tetrahymena*. *Genetics* 70: 251.
- Yao, M.-C., A. Kimmel and M. Gorovsky. 1974. A small number of cistrons for ribosomal RNA in the germinal nucleus of a eukaryote, *Tetrahymena pyriformis*. *Proc. Nat. Acad. Sci. USA*. In press.
- Zellweger, A., U. Ryser and R. Braun. 1972. Ribosomal genes of *Physarum*: their isolation and replication in the mitotic cycle. *J. Mol. Biol.* 64: 681.

Differential replication of rDNA during insect development

J. A. Thomson

The proportion of ribosomal DNA to total DNA from whole insects has been examined at six stages during the development of the brown blowfly, *Calliphora stygia*. Saturation values for the hybridization of rRNA with total DNA shows that DNA from hatching larvae contains approximately 20% more rDNA than does total DNA from either mid-embryonic or adult insects. DNA from late third instar larvae close to pupariation contains about 30% less rDNA than that from hatching larvae. The proportion of rDNA to total DNA in larvae of intermediate ages falls between these extremes. Differentially enhanced replication of rDNA thus takes place in late embryonic or very early larval development in the Diptera, apparently as part of the sequential program of differentiation in cells destined to become highly polytenized.

Growth in the main larval tissues of the higher Diptera involves cell enlargement without mitosis in post-embryonic stages. Many larval tissues in these insects consist of highly specialized, terminally differentiated cells, often with highly polytenized chromosomes. Two such tissues which contribute much to the bulk of the feeding larva are the fat body (Price, 1973; Thomson, 1973a, 1975), which is specialized for the production (Munn et al., 1969) and later accumulation of very large amounts of the storage protein calliphorin (Munn et al., 1967), and the larval mid-gut. Many other less voluminous larval tissues also develop highly polytenized chromosomes, including salivary glands, malpighian tubules, ventral nephrocytes and pericardial cells (see e.g. Thomson and Gunson, 1970), as well as certain more isolated cells associated with the tracheal, nervous and integumental systems. By contrast, few adult cells have polytene chromosomes, and those that do comprise only a small fraction of the bulk of the insect. In *Calliphora*, polytenization is known to occur in the imaginal footpad and trichogen nuclei (Thomson, 1969), in the ovarian nurse cells (Bier, 1957) and in the cells of the malpighian tubules carried over from larval to adult life. Together, these cells make up only a very small fraction of the total cell complement of the adult fly. Thus the composition of total DNA extracts from whole insects should reflect the development of increasing levels of polyteny during larval life, and then

the extent of replacement of larval by adult tissues during metamorphosis.

For some time it has been clear from cytological and cytochemical observations that differential under-replication of rRNA cistrons occurs during polytenization in *Drosophila* (Rudkin, 1972, for review). Hennig and Meer (1971) showed directly that the proportion of rDNA to total DNA decreases during polytenization in the salivary gland of *D. melanogaster*. Sibatani's (1971) finding that the proportion of rDNA is lower in total larval than total adult DNA implies that such differential under-replication of rDNA is both of general occurrence amongst larval tissues, and of sufficient scale to have an overriding influence on the composition of total larval DNA preparations.

An apparent paradox therefore becomes evident in considering the general pattern of larval development. Extraordinarily active protein synthesis occurs in larval tissues such as the fat body (Thomson, 1973a) and the salivary glands of the Diptera. Yet rDNA is under-represented in the polytenized chromosomes of these tissues, just where there would seem to be most requirement for amplification of the rRNA cistrons. The present work was undertaken to examine the possibility that polytenization might be *preceded* by a phase of over-replication of rDNA relative to the remainder of the genome, a situation overtaken and later replaced by relative under-representation of the rDNA as high levels of polyteny are attained (Thomson, 1973b). Such a process would, of course, be consistent with the finding of Spear and Gall (1973) that rDNA is replicated independently of the remainder of the genome in polytene nuclei of larval *Drosophila*.

METHODS

An inbred laboratory strain (MEL) of the Australian brown blowfly, *Calliphora stygia* (Fabr.) was reared at 21° with 16 hour photoperiod. Excess lean mammalian muscle was available to larvae, and cultures were kept uncrowded. Batches of whole insects (5-100 g) were collected at each of the stages indicated in Table 1.

Crude fibrous DNA was obtained from insects at selected ages by the procedure of Hastings and Kirby (1966). This product was purified by enzymatic digestion (Thomson, 1973b), deproteinized and reprecipitated with ethanol.

³²P-rRNA was prepared from actively feeding third instar larvae. These larvae received only as much meat soaked in ³²P-orthophosphate (1.0 mCi/g meat) as they could consume from hatching to harvest 5 days later. RNA, extracted according to Hastings and Kirby (1966),

was further purified by salt extraction of polysaccharides, reprecipitation as the cetyltrimethyl-ammonium salt, and removal of low molecular weight RNA (Thomson, 1973b) before isolation of the rRNA by MAK chromatography. The final product had a specific activity of 15,000 counts/min/ μg .

Replicate nitrocellulose filters, loaded with an estimated 35-50 μg of alkali-denatured-DNA from insects of each batch were incubated with ^{32}P -rRNA at concentrations of 1, 5, 10, 15 and 20 μg rRNA/ml in saline citrate - 40% formamide - 0.05% sodium dodecyl sulphate for 16 hours at 40° . After washing, the filters were treated with ribonuclease, rewashed, dried and counted. Subsequently, the DNA was hydrolyzed from each filter and determined spectrophotometrically at 260 nm. Theoretical saturation values for hybridization of each DNA sample with rRNA were calculated from the reciprocal of the slope of (RNA concentration)/(hybridization) plotted against RNA concentration (Bishop et al., 1969).

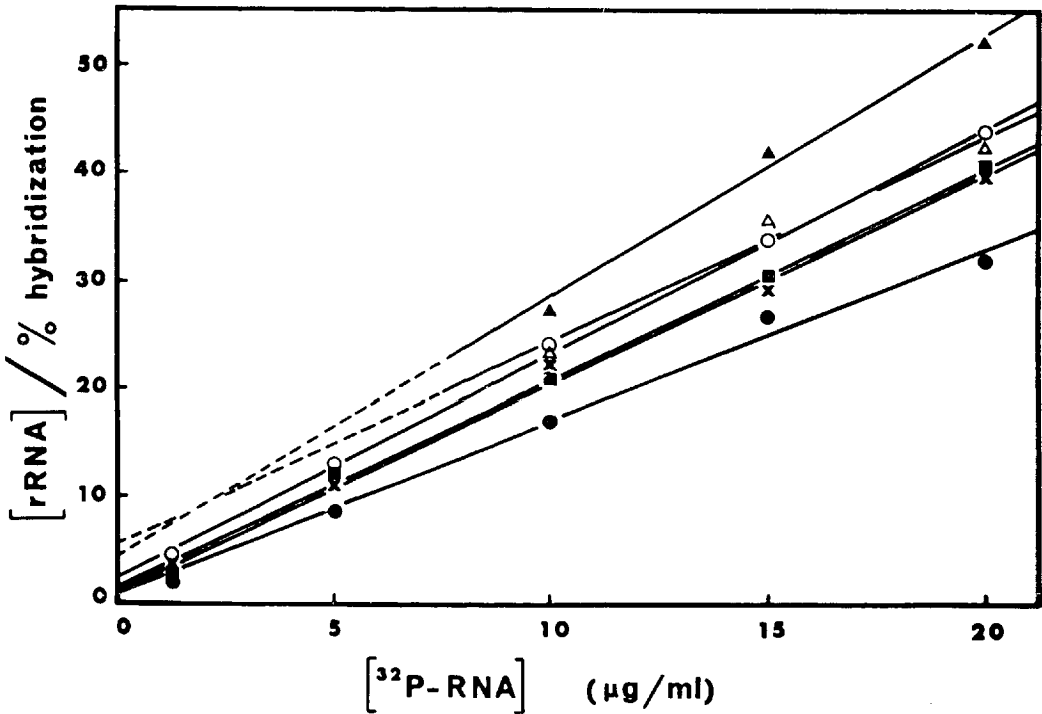


Figure 1. Saturation curves for *Calliphora* DNA incubated with rRNA, shown as the ratio [concentration of ^{32}P -rRNA]/[percentage hybridization] plotted against concentration of ^{32}P -rRNA. Each point represents the mean of 2-4 replicates, agreeing to within 5%.
 ○, Embryo; ●, first instar; ×, second instar; △, mid-third instar; ▲, late third instar; ■, adult.

RESULTS AND DISCUSSION

The calculated saturation value (r_s) for mid-embryonic DNA (0.48, Table 1) is close to that for adult DNA (0.51). The slopes for the reciprocal plots of hybridization for these two samples at various ^{32}P -rRNA concentrations (Figure 1) do not differ significantly (t-test). However, DNA from first instar larvae differs significantly from samples of DNA from embryos ($P < 0.001$), late third instar larvae ($P < 0.01$) and adults ($P < 0.01$). Major changes in the proportion of rDNA to total DNA in larval development thus involve first an increase of 20% above the mid-embryonic level, then major decreases in the second and late third instars (Table 1). By the end of larval life total DNA contains 30% less rDNA than that from hatching larvae.

TABLE 1

Calculated saturation values (r_s) for hybridization of ^{32}P -rRNA with DNA extracted from *Calliphora* at six developmental stages

Source of DNA	r_s
Embryo	
12 \pm 3 hr ^a	0.48
Larva	
Instar 1, 0.15-0.20 mg, 24-26 hr ^a	0.63
Instar 2, 2.5 \pm 0.5 mg, 2 days ^b	0.52
Instar 3, 100 \pm 10 mg, 5-6 days ^b	0.53
Instar 3, 100 \pm 10 mg, 10-11 days ^b	0.41
Adult	
2-6 hr ^c , unfed, mixed sex	0.51

The formation and fate of the nucleoli can easily be followed in the larval tissues of *Calliphora* (Thomson, 1973a, 1975; Thomson and Gunson, 1970). A single nucleolus is found at hatching in the range of tissues examined. In a period of rapid enlargement, the nucleolus opens out into a horseshoe or ring shape. By mid-second instar the nucleolus fragments to form a tissue-specific pattern of prominent ribonucleoprotein masses. These become chromosomally

associated during the main period of protein synthesis, which extends in the tissues of feeding larvae from mid-second instar into the first one-third of the third instar. Development of large multiple nucleolar masses precedes the principal phase of protein synthesis but continues during its early stages, e.g., in both fat body and salivary gland (Thomson, 1973a). After the main burst of larval protein synthesis subsides, only remnants of the nucleolar masses remain until a secondary nucleolar body is reconstituted shortly before the end of larval life (Thomson, 1975; Thomson and Gunson, 1970). The replication cycles involved in attainment of high levels of polyteny occur subsequent to establishment of the multiple nucleolar system. Chromosomal replication thus continues after the nucleolar events correlated with peak protein synthesis have been completed.

In *Calliphora*, then, cytological observations of nucleolus formation and of polytenization, the pattern of protein synthesis in larval tissues (Kinnear et al., 1971; Thomson, 1975) and the changes in the relative rDNA content of whole insects are consistent with a developmental programme in which differentially enhanced rDNA synthesis precedes polytenization. This sequence permits rRNA synthesis to start before the main phase of transcription, processing and transport of the mRNAs involved in the large scale synthesis of the specialized proteins of the larval tissues.

REFERENCES

- Bier, K. 1957. Endomitose und Polytenie in den Nahrzellkernen von *Calliphora erythrocephala* Meigen. *Chromosoma* 10: 493.
- Bishop, J.O., F.W. Robertson, J.A. Burns and M. Melli. 1969. Methods for the analysis of deoxyribonucleic acid-ribonucleic acid hybridization data. *Biochem. J.* 115: 361.
- Hastings, J.R.B. and K.S. Kirby. 1966. The nucleic acids of *Drosophila melanogaster*. *Biochem. J.* 100: 532.
- Hennig, W. and B. Meer. 1971. Reduced polyteny of ribosomal RNA cistrons in giant chromosomes of *Drosophila hydei*. *Nature New Biol.* 233: 70.
- Kinnear, J.F., M.D. Martin and J.A. Thomson. 1971. Developmental changes in the late larva of *Calliphora stygia*. III. The occurrence and synthesis of specific tissue proteins. *Aust. J. Biol. Sci.* 24: 275.
- Munn, E.A., A. Feinstein and G.D. Greville. 1967. A major protein constituent of pupae of the blowfly, *Calliphora erythrocephala* (Diptera). *Biochem. J.* 102: 5.
- Munn, E.A., G.M. Price and G.D. Greville. 1969. The synthesis *in vivo* of the protein calliphorin by fat body from the blowfly, *Calliphora erythrocephala*. *J. Insect Physiol.* 15: 1601.

- Price, G.M. 1973. Protein and nucleic acid metabolism in insect fat body. *Biol. Reviews* 48: 333.
- Rudkin, G.T. 1972. Replication in polytene chromosomes. In Results and Problems in Cell Differentiation (W. Beermann, ed.) Vol. 4, p.59. Springer-Verlag, Berlin.
- Sibatani, A. 1971. Difference in the proportion of DNA specific to ribosomal RNA between adults and larvae of *Drosophila melanogaster*. *Mol. Gen. Genet.* 114: 177.
- Spear, B.B. and J.G. Gall. 1973. Independent control of ribosomal gene replication in polytene chromosomes of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 70: 1359.
- Thomson, J.A. 1969. The interpretation of puff patterns in polytene chromosomes. *Currents Modern Biology* 2: 333.
- _____. 1973a. Patterns of gene activity in larval tissues of the blowfly, *Calliphora*. In The Biochemistry of Gene Expression in Higher Organisms (J.K. Pollak and J.W. Lee, ed.) p.320. Australia and New Zealand Book Company, Sydney.
- _____. 1973b. Differential replication of ribosomal DNA during larval development in *Calliphora* (Diptera). *Dev. Biol.* 35: 362.
- _____. 1975. Major patterns of gene activity during development in holometabolous insects. *Adv. Insect Physiol.* 11: In press.
- Thomson, J.A. and M.M. Gunson. 1970. Developmental changes in the major inclusion bodies of polytene nuclei from larval tissues of the blowfly, *Calliphora stygia*. *Chromosoma* 30: 193.

In vivo purification of mitochondrial genes in petite mutants of yeast: the genetic and biochemical characterization of mutants enriched in mitochondrial ribosomal RNA cistrons

Phillip Nagley, P. L. Molloy, H. B. Lukins and Anthony W. Linnane

Cytoplasmic petite mutants of yeast contain fragments of the mitochondrial genome which, in different clones, vary as to the information content of the segment of DNA retained. A cellular regulatory mechanism maintains the total amount of mitochondrial DNA at a fixed and high level in different petites irrespective of the sequences in the mitochondrial DNA. This system of gene purification has been studied by isolating petite mutants retaining one of four regions in mitochondrial DNA genetically defined by the presence of specific antibiotic resistance loci, and testing the petite mitochondrial DNAs for the ability to hybridize with mitochondrial ribosomal RNA. The results indicate that the larger (21S) ribosomal RNA cistron is located close to genetic loci for erythromycin and chloramphenicol resistance, and distant from loci for oligomycin and mikamycin resistance. One petite carrying only the erythromycin resistance locus contains mitochondrial DNA of which more than one quarter is a sequence complementary to 21S ribosomal RNA.

There has been considerable interest in recent years in the concept of purification of specific small regions of relatively large genomes for biochemical and molecular analysis. Attention has turned to the isolation of segments of larger genomes which do not separate readily from the bulk of the DNA by virtue of specific physical or biochemical properties, as do satellite DNAs and some organelle DNA molecules. An example of the purification of a DNA segment otherwise buried within a bacterial genome is the purification of the *lac* operon genes of *Escherichia coli* attached to defective bacteriophage DNA (Reznikoff, 1972). In the case of eukaryote genomes, randomly produced fragments of *Drosophila* DNA have recently been joined *in vitro* to a bacterial plasmid DNA which is then reintroduced into *E. coli* so as to permit cloning and reproduction of the eukaryote DNA fragment in amounts able to be physically studied (Hogness, this volume).

In this communication we describe the unique system we have been investigating for several years (see Linnane et al., 1972) which enables the isolation in good yield of selected segments of the mitochondrial DNA (mtDNA) of yeast which have been biologically

purified and amplified in cytoplasmic petite mutants. First we discuss the general properties of this system which enables the isolation of selected genetically defined regions in mtDNA. We then consider our recent results on the characterization of mutants highly enriched for portions of the cistrons specifying ribosomal RNA (rDNA). Finally we consider some general problems related to the phenomenon of mitochondrial gene amplification in petite mutants.

Petite mutants are deletion mutants of the mitochondrial genome
Respiratory competent (grande or ρ^+) haploid yeast cells contain of the order of 50 copies of a mitochondrial genome 50×10^6 daltons in size (see Williamson, 1970; Nagley and Linnane, 1972). Cytoplasmically inherited respiratory deficient mutants (petite) are of two broadly defined types; those petites retaining some mtDNA (ρ^-) and those from which mtDNA has been entirely lost (ρ^0) (Nagley and Linnane, 1970, 1972; Goldring et al., 1970). The mtDNA in ρ^- cells has been shown by molecular hybridization studies using ρ^+ mtDNA, as well as by measurements of the re-association kinetics of denatured ρ^- mtDNA, to have arisen by deletion of different regions of grande mtDNA (reviewed by Borst, 1972; Nagley and Linnane, 1974). Previous genetic studies had already suggested this conclusion (Gingold et al., 1969; Saunders et al., 1971) on the basis of the following observations. Petite mutants derived from a grande strain carrying an antibiotic resistance marker in its mtDNA (e.g. *ery 1-r*) can be shown by a simple genetic test either to retain this locus in mtDNA ($\rho^- \text{ ery } 1-r$) or to have this region of mtDNA deleted ($\rho^- \text{ ery } 1-o$). For this genetic test the petite in question is crossed with a haploid grande strain sensitive to the antibiotic whose locus is under examination. If among grande diploids produced in this cross cells resistant to the antibiotic are found then the resistance locus must have been retained in the haploid petite cells. If the diploids are entirely sensitive to the antibiotic the deleted state (o) is assigned to the petite as regards this resistance locus. As will become evident below, the retention of different regions of mtDNA appears to be largely random and of widely varying extent. Indeed, no single area in mtDNA has thus far been shown to be common to all petites (see Sanders et al., 1973).

The cellular level of mtDNA in ρ^- cells is regulated independently of the segment of the genome retained In 1971 we approached the question as to whether in petite mutants which had suffered different extents of deletion of grande base sequences from the mtDNA, there were corresponding reductions in the cellular level of mtDNA. One might expect *a priori* that petite cells retaining a very small segment of the mitochondrial genome would have a very low cellular

level of mtDNA. That this in fact is not the case was demonstrated by our analysis of many petite mutants derived from a single grande strain (Nagley and Linnane, 1972). The results showed clearly that the cellular level of mtDNA was relatively constant over a wide range of characteristics of mtDNA in the petite clones. Thus, the parent grande strain L411 used in these studies contained 18% of its cellular DNA as mtDNA, and the petite clones derived from it all fell in the range of mtDNA comprising $16 \pm 5\%$ of total DNA. Individual petite clones do show a characteristic level within this range. We have recently extended this study to petite clones derived from another grande strain (L2200) whose cellular level of mtDNA is 24% of total DNA. The petite (*rho*⁻) clones derived from this strain contained a mtDNA level in the range $22 \pm 5\%$ of total DNA (Hall, Nagley and Linnane, unpublished). It is likely that this regulation system in petites is maintained by gene products of nuclear DNA in view of the apparently nonfunctional nature of the petite mitochondrial genome.

The petite genome can thus be envisaged as a deletion of genetic information from the grande genome with a compensatory amplification of the remaining segments to make up the required cellular level of mtDNA. Evidence for this view is presented below. The exact way in which the amplification might occur is also discussed.

Isolation of stable petite clones retaining defined regions of the mitochondrial genome The ability to select for petite clones containing specific segments of mtDNA has been made possible by the identification of several antibiotic resistance loci in mtDNA (see Linnane et al., 1972, for review). The loci studied in this laboratory have been assigned relative positions on a linear map by genetic techniques; the behaviour of the various markers in crosses as regards transmission and recombination frequencies has been analysed (see Linnane et al., 1974). Four main regions have been studied in our laboratory. Representative markers from the clusters of very closely linked loci in each region are *cap 1*, *ery 1*, *oli 1* and *mik 1*. The order of *cap 1* - *ery 1* - *oli 1* has been unambiguously assigned (Linnane et al., 1974; cf. Avner et al., 1973). In the case of *mik 1*, it is known that this locus is relatively far from the *cap 1* and *ery 1* regions but it has not been possible to assign an exact position relative to *oli 1* (Howell et al., 1974).

The most convenient procedure used to obtain petites retaining defined segments of mtDNA requires the construction of grande haploid strains retaining three or more drug resistance loci and the subsequent selection of petite mutants (spontaneously occurring or actively induced by mutagens) retaining or losing the various drug

resistance markers. The isolation of stable petite clones often requires many steps of subcloning to obtain a petite carrying the genotype of interest in a homogeneous and stable population (Saunders et al., 1971; Nagley and Linnane, 1972; Nagley et al., 1974a). In this manner we have isolated a number of petite clones carrying one or two specific loci with other loci demonstrably eliminated. Two extra points should be made. Firstly, although it is convenient to select for petites retaining antibiotic resistance loci, it is possible to test a petite so selected for the retention of any other locus which was initially present as the sensitive allele in the parent grande strain. For this test, the petite is crossed with an appropriate grande resistant strain and the grande diploids resulting from the cross are tested for resistance or sensitivity to the particular antibiotic (see Saunders et al., 1971; Nagley et al., 1974b). The presence of sensitive grande diploids indicates that the petite mitochondrial genome contains the sensitive allele of that locus, which is introduced into the grande diploids presumably by recombination between the petite and grande mtDNA molecules (cf. Gingold et al., 1969). The second point to note is that when a random population of petites derived from a suitable grande strain carrying several resistance loci is examined for the frequency of co-retention of the various markers in individual petite clones, further information on the proximity of loci can be obtained. From such an analysis (Molloy, Lukins and Linnane, unpublished) it was found that the *cap 1* and *ery 1* loci are either retained or lost together in the great majority of cases. This tight linkage suggests close physical proximity (cf. Suda and Uchida, 1974). Clones in which either *cap 1* or *ery 1* loci are retained individually in the absence of the other are quite rare (cf. Faye et al., 1973). By contrast, *oli 1* and *mik 1* loci separate readily from each other and also separate easily from the *cap 1* and *ery 1* regions.

Characterization of petite clones retaining ribosomal RNA genes The two RNA molecules (15S and 21S) found in mitochondrial ribosomes (Grivell et al., 1971; Morimoto et al., 1971) are transcribed from genes in *rho*⁺ mtDNA, there being one cistron for each per mitochondrial genome of 50×10^6 daltons (see below). These rRNA-cistrons can be regarded as a specific segment of mtDNA to be retained or lost in different petite clones according to the concepts presented above. It is not possible to select directly for petite clones retaining the rRNA genes in the manner described for selection of drug resistance locus retention. It is, however, feasible to determine the retention or loss of rDNA in petites retaining various genetically defined segments of mtDNA. From such an approach would derive information as to the location of the rRNA cistrons in mtDNA

relative to the various drug resistance loci. More importantly, a direct test of the gene purification scheme can be made. This scheme predicts that it should be possible to isolate petite clones in which the mtDNA segment retained is enriched for the sequence homologous to rRNA. In such clones deletion of extraneous mtDNA segments would be expected to increase the contribution of rDNA sequences to the total genetic information present in mtDNA.

In order to carry out these analyses we developed a convenient procedure for measuring the hybridization of mitochondrial rRNA to mtDNA. The experimental difficulties to be overcome primarily related to the yield of mitochondrial rRNA, the purity of rRNA species and the specific radioactivities needed for the detection of hybrids with mtDNA. The procedures adopted have been detailed by Nagley et al. (1974b) and involved an initial preparation and separation of the mitoribosome subunits (37S and 50S), from grande mitochondria. The subunits were used as the source of each rRNA species; this was necessary to avoid the contamination of rRNA with mRNA, which has previously been shown to cause serious problems in RNA-DNA hybridization experiments where the rRNAs had been prepared by sucrose gradient fractionation of whole mitochondrial RNA (Reijnders et al., 1972; Schafer and Kuntzel, 1972). After extraction from subunits, each purified rRNA was labelled by iodination *in vitro* with (^{125}I) to specific activities of about 2×10^5 cpm/ μg RNA (Getz et al., 1972). The hybridization of labelled 15S and 21S rRNA with a preparation of a grande mtDNA is shown in Figure 1, panel A. The slope of the linear part of the graphs where RNA is bound in exact proportion to the quantity of DNA added is used to calculate the fraction of grande mtDNA homologous to each rRNA. The observed deviation from linearity at higher DNA inputs is considered to arise from reassociation between DNA molecules competing with the RNA hybridizing to DNA, thus preventing quantitative hybrid formation. From the slopes of the linear regions [^{125}I] cpm bound/ μg DNA], and the measured specific activities of each labelled RNA species, it is calculated that 0.7% of grande mtDNA is homologous to 15S rRNA, and 1.5% to 21S rRNA. These values agree almost exactly with the saturation plateaux measured in Borst's and in Halvorson's laboratories where denatured mtDNA was bound to filters and incubated in solutions containing various amounts of (^{32}P) labelled mitochondrial rRNA (Reijnders et al., 1972; Morimoto et al., 1971). These saturation levels are consistent with one cistron each for 15S rRNA (size 0.6×10^6 daltons) and 21S rRNA (size 1.6×10^6 daltons); the observed saturation levels are about 30% lower than expected on theoretical grounds for one cistron each (see Reijnders et al., 1972, for discussion of this point).

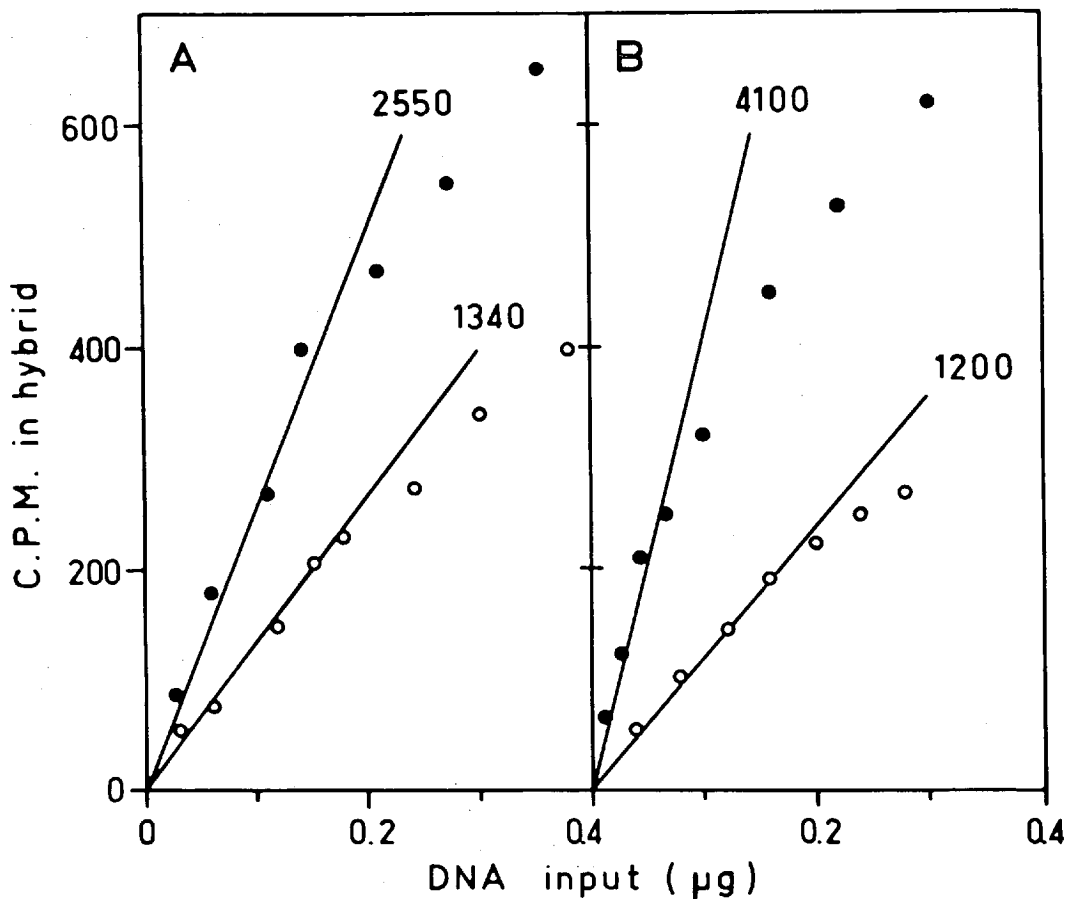


Figure 1. Hybridization of mitochondrial rRNA with mtDNA from (A) strain L411 [ρ^+ *ery 1-r*] and (B) petite clone KIE13-6 [ρ^- *ery 1-r cap 1-o oli 1-o mik-1-o*] derived from strain L411 by ethidium bromide mutagenesis and extensive subcloning. Details of nucleic acid preparations and hybridization procedures have been given elsewhere (Nagley et al., 1974b). For hybridization, a fixed amount (0.12 μg) of labelled RNA either 15S rRNA (o) 1.92×10^5 cpm/ μg ; or 21S rRNA (●) 1.65×10^5 cpm/ μg was incubated with different amounts of denatured mtDNA in a solvent containing 30% formamide and 0.33 M Na^+ (2 x SSC) at 37° for 2 hours. After RNase digestion, hybrids were collected on nitrocellulose filters, which were counted for gamma radioactivity. The gradient of the linear portion of each curve (cpm/ μg DNA) is indicated on the figure.

The data we have obtained for several petite clones retaining one drug resistance marker is summarised in Table 1. Details for most of the hybridization reactions have been presented elsewhere (Nagley et al., 1974b), except for petite KIE13-6 (Figure 1,

Purification of mitochondrial genes 161

TABLE 1

Characteristics of mitochondrial genomes of a grande strain and petite clones retaining a single antibiotic resistance locus

Character	Grande	Petite clones				
	L411	P2	EP1	KIE13-6	S2	S1
Locus retained ^a	(all)	<i>cap 1</i>	<i>ery 1</i>	<i>ery 1</i>	<i>oli 1</i>	<i>mik 1</i>
Cellular level of mtDNA (% total DNA) ^b	18	11	10	15	17	11
Density of mtDNA in CsCl (g/ml)	1.683	1.681	1.687	1.683	1.681	1.683
Homology to rRNA (% DNA) ^c						
15S	0.7	n.s. ^d	n.s.	0.6	n.d. ^e	n.d.
21S	1.5	1.9	14.0	2.5	n.d.	0.3
Enrichment of rDNA over grande ^f						
15S	1	0	0	0.9	0	0
21S	1	1.3	9.3	1.7	0	0.2

^a Each petite clone was purified for the indicated mitochondrial genotype by extensive subcloning. For each petite clone, the presence of the three loci other than the locus retained was checked (including sensitive alleles: see text); these were shown to be deleted.

^b Cellular DNA was extracted, and analysed in CsCl gradients for the percentage of cellular DNA which is mtDNA, as described by Nagley and Linnane (1972).

^c Data for L411 and KIE13-6 are calculated from Figure 1; remaining data are taken from Nagley et al. (1974b).

^d n.s. (not significant) indicates that although a small degree of hybridization was observed (up to 100 cpm) with 15S rRNA, competition tests with unlabelled 15S rRNA showed that the counts bound to DNA were not in fact due to label in authentic 15S rRNA, but probably to minor contaminating labelled material (Nagley et al., 1974b). Whereas unlabelled 15S rRNA strongly competed hybridization of labelled 15S rRNA with mtDNA of L411 and KIE13-6, no such competition was observed when mtDNA of P2 and EP1 was used. It was instead found that unlabelled 21S rRNA competed very strongly with the hybridization of labelled 15S rRNA to mtDNA of P2 and EP1, but not to L411 and KIE13-6.

^e n.d. indicates no detectable counts in hybrids were found.

^f Enrichment is defined as the percentage homology of a rRNA with a petite mtDNA divided by the percentage homology of that rRNA with grande mtDNA.

panel B). Considering the data in Table 1, it can be seen that petites S1 and S2 respectively retaining *mik 1* and *oli 1* loci alone show very little or no hybridization with either 15S or 21S rRNA; the mitochondrial rDNA is thus largely if not completely deleted from these two clones. Where *cap 1* or *ery 1* loci are retained, substantial hybridization with 21S rRNA is observed. It is inferred that the 21S rRNA gene is located close to *cap 1* and *ery 1* loci in grande mtDNA. In particular, clone EP1 is highly enriched for DNA homologous to 21S rRNA. This clone contains mtDNA of which 14% hybridizes to 21S rRNA, indicating that more than one quarter of the information retained in its mtDNA is rDNA. Among several possibilities to explain this high enrichment for rDNA two important ones are to be considered. Firstly, it is clear that considerable deletion of non-rDNA has occurred in clone EP1 because of the observed loss of three antibiotic resistance loci. It is not certain whether simple deletion is the only explanation for the enrichment of rDNA in the segment of mtDNA retained. Thus, Rabinowitz and colleagues have studied mtDNA in a petite clone in which the leucyl-tRNA gene is enriched six-fold in the petite mtDNA compared with grande mtDNA. Surprisingly, the mtDNA segment retained in the petite was homologous to about 70% of the grande mtDNA, suggesting that simple deletion was insufficient to explain the six-fold enrichment (Gordon et al., 1974; Casey et al., 1974). It is thus conceivable that differential amplification of various segments of mtDNA present in one petite clone can take place; this is supported by the bimodal renaturation kinetics of some petite mtDNAs (Casey et al., 1974). We cannot yet describe the nature of the enrichment of rDNA in clone EP1 until we have measured the fraction of the grande genome to which EP1 mtDNA is homologous as well as the fraction of the 21S gene which is retained in clone EP1. These determinations are currently in progress.

It can be noted that in the majority of petite clones in Table 1, no detectable hybridization with 15S rRNA was observed. However, analysis of petite clone KIE13-6, which retains the *ery 1* locus alone shows substantial hybridization with both 15S and 21S rRNA species (Figure 1, panel B). The relative enrichments of 15S and 21S genes are different in petite clone KIE13-6; the relative hybridization levels of 15S and 21S rRNA are no longer in the ratio 1:2 as they are in grande mtDNA. It has not been established yet if this situation in KIE13-6 arises because the fractional length of the 21S gene is greater than fractional length of the 15S gene retained, or alternatively, whether the portion of 21S gene retained is differentially amplified with respect to that portion of the 15S gene.

Buoyant density changes in petite mtDNA In a number of petite clones listed in Table 1, the buoyant density of mtDNA is different from that of the grande (1.683 g/ml). This change in density in petite clones has been reported many times before (Mounolou et al., 1966; reviewed by Borst, 1972, Nagley and Linnane, 1974) and can now be interpreted to indicate the retention of a region of mtDNA biased towards a base composition different from the average grande mtDNA base composition (17% G+C). Segments of A+T-rich sequences of lengths between 10^5 and 10^6 daltons have been shown to be present in grande mtDNA (see Bernardi et al., 1972; Piperno et al., 1972). The occurrence of these discrete types of stretches may lead to the anomalous densities in CsCl of yeast mtDNAs relative to the average base compositions. Most reported density changes in petite mtDNA have been in the direction of a reduction in density (i.e., higher A+T content), and changes to elevated density (increased G+C content) are relatively rare. Indeed such changes were only unequivocally identified by selection of petites purified for the segment containing the *ery 1* locus, exemplified by clone EPl in Table 1. The high enrichment of this genome for 21S rRNA sequences (G+C content 26%) is consistent with the observed shift to greater density of the mtDNA as a whole. A series of petite clones purified for *ery 1* loci have been described by Michaelis et al. (1973) and these also show densities of mtDNA of about 1.687 g/ml. It is important to note that different petite clones retaining the same detected loci do not always show the same density in mtDNA. Thus clone KIE13-6 has a density of 1.683 g/ml, although it shares with EPl the property of only retaining *ery 1* loci.

On the nature of the amplification process of the segments of mtDNA retained in petite cells From Table 1 it will be noted that all the petite clones studied here follow the principle discussed above that their cellular level of mtDNA is maintained at a high level irrespective of the nature of the information retained in the mtDNA. All petites in Table 1 were derived from grande strains which contain 18% of their cellular DNA as mtDNA, and all petites fall in the range of 10-17% mtDNA. That true amplification of the segment of mtDNA retained in petite mutants occurs rather than replacement of deleted sequences by some kind of 'non-informational' DNA which is synthesised to build up the level of mtDNA is supported by the following independent lines of evidence. (1) In the [*rho*⁻ *ery 1-r*] cells described by Nagley et al. (1973, 1974a), it was shown by a target analysis procedure using ethidium bromide as mutagen that there were indeed more copies of *ery 1-r* genes per petite cell than per grande cell as predicted if amplification does occur. (2) The data presented in this paper show the enrichment of mtDNA for rDNA regions in certain petites which have lost other genetically defined

regions of mtDNA. (3) Sequence homology measurements between grande and petite mtDNA have shown that petites contain DNA complementary to only a portion of the grande mtDNA, but that petites rarely are found to contain sequences in petite mtDNA unable to hybridize to grande mtDNA (i.e. new sequences) (Michaelis et al., 1972; Gordon et al., 1974). In one case a small portion of petite mtDNA (15%) has been found unable to hybridize to grande mtDNA (Gordon and Rabinowitz, 1973).

The general nature of amplification has not yet been extensively considered in molecular terms. One model that can be proposed is that the reduction in genome size in a petite is not in fact accompanied by a reduction in the physical size of mtDNA inside the cells. The regulation of mtDNA levels may be at the level of regulating the total number of molecules of mtDNA per cell, and variations in mtDNA levels in different petites may relate to the length of individual DNA molecules in the cells. It is thus envisaged that amplification may occur by a process of end-to-end reduplication of the mtDNA segment retained in a petite cell to give long linear mtDNA molecules containing tandem repeating units. Indeed recent electron microscopic examination of partially denatured mtDNA from several petite clones has strongly suggested the existence of repeated segments within linear DNA molecules (Faye et al., 1973). Some of these may be tandem inverted repeats (Rabinowitz et al., 1974) as well as head-to-tail regular repeated segments.

Very small circles have been observed in petite mtDNA preparations (Hollenberg et al., 1972; Clark-Walker and Miklos, 1973). The physical sizes of such circles have been shown to be equivalent to the lengths of the segments of mtDNA retained in a series of petite clones as measured by renaturation kinetics (Rabinowitz et al., 1974). It is important to learn whether these small molecules represent *precursors* of the long molecules which form by far the bulk of mtDNA (e.g. by rolling circle mechanisms) or whether these small circles represent excision *products* from the long molecules by some process of internal recombination, which for example may occur during replication of the long molecules containing tandem repeats and which thus have many regions of internal homology.

REFERENCES

- Avner, P.R., D. Coen, B. Dujon and P.P. Slonimski. 1973. Mitochondrial genetics. IV. Allelism and mapping studies of oligomycin resistant mutants in *S. cerevisiae*. *Mol. Gen. Genet.* 125: 9.
- Bernardi, G., G. Piperno and G. Fonty. 1972. The mitochondrial genome of wild type yeast cells. I. Preparation and heterogeneity of mitochondrial DNA. *J. Mol. Biol.* 65: 173.

- Borst, P. 1972. Mitochondrial nucleic acids. *Ann. Rev. Biochem.* 41: 333.
- Casey, J., P. Gordon and M. Rabinowitz. 1974. Characterization of mitochondrial deoxyribonucleic acids from grande and petite yeasts by renaturation and denaturation analysis and by transfer ribonucleic acid hybridization: Evidence for repetition or heterogeneity in mitochondrial deoxyribonucleic acid populations. *Biochemistry* 13: 1059.
- Clark-Walker, G.D. and G.L.G. Miklos. 1974. Localization and quantification of circular DNA in yeast. *Eur. J. Biochem.* 41: 359.
- Faye, G., H. Fukuhara, C. Grandchamp, J. Lazowska, F. Michel, J. Casey, G.S. Getz, J. Locker, M. Rabinowitz, M. Bolotin-Fukuhara, D. Coen, J. Deutsch, B. Dujon, P. Netter and P.P. Slonimski. 1973. Mitochondrial nucleic acids in the petite colonie mutants: Deletions and repetitions of genes. *Biochimie* 55: 779.
- Getz, M.J., L.C. Altenburg and G.F. Saunders. 1972. The use of RNA labelled *in vitro* with iodine-125 in molecular hybridization experiments. *Biochim. Biophys. Acta* 287: 485.
- Gingold, E.B., G.W. Saunders, H.B. Lukins and A.W. Linnane. 1969. Biogenesis of mitochondria. X. Reassortment of the cytoplasmic genetic determinants for respiratory competence and erythromycin resistance in *Saccharomyces cerevisiae*. *Genetics* 62: 735.
- Goldring, E.S., L.I. Grossman, D. Krupnick, D.R. Cryer and J. Marmur. 1970. The petite mutation in yeast. I. The loss of mitochondrial DNA during petite induction with ethidium bromide. *J. Mol. Biol.* 52: 323.
- Gordon, P., J. Casey and M. Rabinowitz. 1974. Characterization of mitochondrial deoxyribonucleic acid from a series of petite strains by deoxyribonucleic acid-deoxyribonucleic acid hybridization. *Biochemistry* 13: 1067.
- Gordon, P. and M. Rabinowitz. 1973. Evidence for deletion and changed sequence in the mitochondrial deoxyribonucleic acid of a spontaneously generated petite mutant of *Saccharomyces cerevisiae*. *Biochemistry* 12: 116.
- Grivell, L.A., L. Reijnders and P. Borst. 1971. Isolation of yeast mitochondrial ribosomes highly active in protein synthesis. *Biochim. Biophys. Acta* 247: 91.
- Hollenberg, C.P., P. Borst and E.F.J. Van Bruggen. 1972. Mitochondrial DNA from cytoplasmic petite mutants of yeast. *Biochim. Biophys. Acta* 277: 35.

- Howell, N., P.L. Molloy, A.W. Linnane and H.B. Lukins. 1974. Biogenesis of mitochondria. 34. The synergistic interaction of nuclear and mitochondrial mutations to produce resistance to high levels of mikamycin in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 128: 43.
- Linnane, A.W., J.N. Haslam, H.B. Lukins and P. Nagley. 1972. The biogenesis of mitochondria in micro-organisms. *Ann. Rev. Microbiol.* 26: 163.
- Linnane, A.W., N. Howell and H.B. Lukins. 1974. Mitochondrial genetics. In *The Biogenesis of Mitochondria: Transcriptional, Translational and Genetic Aspects* (A.M. Kroon and C. Saccone, ed.) p.193. Academic Press, New York.
- Michaelis, G., S. Douglass, M. Tsai, K. Burchiel and R.S. Criddle. 1972. *In vitro* transcription of mitochondrial deoxyribonucleic acid from yeast. *Biochemistry* 11: 2026.
- Michaelis, G., E. Petrochilo and P.P. Slonimski. 1973. Mitochondrial genetics. III. Recombined molecules of mitochondrial DNA obtained from crosses between cytoplasmic petite mutants of *Saccharomyces cerevisiae*: Physical and genetic characterization. *Mol. Gen. Genet.* 123: 51.
- Morimoto, H., A.H. Scragg, J. Nekhorocheff, V. Villa and H.O. Halvorson. 1971. Comparison of the protein synthesising systems from mitochondria and cytoplasm of yeast. In *Autonomy and Biogenesis of Mitochondria and Chloroplasts* (N.K. Boardman, A.W. Linnane and R.M. Smillie, ed.) p.282. North Holland, Amsterdam.
- Mounolou, J.C., H. Jakob and P.P. Slonimski. 1966. Mitochondrial DNA from yeast 'petite' mutants: Specific changes of buoyant density corresponding to different cytoplasmic mutations. *Biochem. Biophys. Res. Comm.* 24: 218.
- Nagley, P., E.B. Gingold and A.W. Linnane. 1974a. The purification of mitochondrial genes using petite mutants of yeast. In *The Biogenesis of Mitochondria: Transcriptional, Translational and Genetic Aspects* (A.M. Kroon and C. Saccone, ed.) p.157. Academic Press, New York.
- Nagley, P., E.B. Gingold, H.B. Lukins and A.W. Linnane. 1973. Biogenesis of mitochondria. XXV. Studies on the mitochondrial genomes of petite mutants of yeast using ethidium bromide as a probe. *J. Mol. Biol.* 78: 335.
- Nagley, P. and A.W. Linnane. 1970. Mitochondrial DNA deficient petite mutants of yeast. *Biochem. Biophys. Res. Comm.* 39: 989.
- , ———. 1972. Biogenesis of mitochondria. XXI. Studies on the nature of the mitochondrial genome in yeast: The degenerative effects of ethidium bromide on mitochondrial genetic information in a respiratory competent strain. *J. Mol. Biol.* 66: 181.

- , ———. 1974. Mitochondrial DNA. *In* Advances in Molecular Genetics (I.R. Falconer and D. Cove, ed.). Paul Elek, London.
- Nagley, P., P.L. Molloy, H.B. Lukins and A.W. Linnane. 1974b. Studies on mitochondrial gene purification using petite mutants of yeast: Characterization of mutants enriched in mitochondrial ribosomal RNA cistrons. *Biochem. Biophys. Res. Comm.* 57: 232.
- Piperno, G., G. Fonty and G. Bernardi. 1972. The mitochondrial genome of wild type yeast cells. II. Investigations on the compositional heterogeneity of mitochondrial DNA. *J. Mol. Biol.* 65: 191.
- Rabinowitz, M., J. Casey, P. Gordon, J. Locker, H. Hsu and G.S. Getz. 1974. Characterization of yeast grande and petite mitochondrial DNA by hybridization and physical techniques. *In* The Biogenesis of Mitochondria: Transcriptional, Translational and Genetic Aspects (A.M. Kroon and C. Saccone, ed.) p.89. Academic Press, New York.
- Reijnders, L., C.M. Kleisen, L.A. Grivell and P. Borst. 1972. Hybridization studies with yeast mitochondrial RNAs. *Biochim. Biophys. Acta* 272: 396.
- Reznikoff, W.S. 1972. The operon revisited. *Ann. Rev. Genet.* 6: 133.
- Sanders, J.P.M., R.A. Flavell, P. Borst and J.N.M. Mol. 1973. Nature of the base sequence conserved in the mitochondrial DNA of a low density petite. *Biochim. Biophys. Acta* 312: 441.
- Saunders, G.W., E.B. Gingold, M.K. Trembath, H.B. Lukins and A.W. Linnane. 1971. Mitochondrial genetics in yeast: Segregation of a cytoplasmic determinant in crosses, and its loss or retention in the petite. *In* Autonomy and Biogenesis of Mitochondria and Chloroplasts (N.K. Boardman, A.W. Linnane and R.M. Smillie, ed.) p.185 North Holland, Amsterdam.
- Schafer, K.P. and H. Kuntzel. 1972. Mitochondrial genes in Neurospora: A single cistron for ribosomal RNA. *Biochem. Biophys. Res. Comm.* 46: 1312.
- Suda, K. and A. Uchida. 1974. The linkage relationship of the cytoplasmic drug-resistance factors in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 128: 331.
- Williamson, D.H. 1970. The effect of environmental and genetic factors on the replication of mitochondrial DNA in yeast. *In* Control of Organelle Development (P.L. Miller, ed.) p.247. University Press, Cambridge.

Genes and chromomeres of *Drosophila*

Burke H. Judd

Evidence which indicates that a basic unit of genetic function in *Drosophila* occupies a single band or chromomere of the polytene chromosomes, opens several important questions about the organization of genes in this eukaryote. One of these questions is whether the gene as identified by the usual operational criteria of mutation, recombination and function, occupies all of the DNA of a chromomere or only a small portion of it. The size range of chromomeres is from less than 5,000 to more than 100,000 nucleotide base pairs per haploid strand equivalent. At the lower end of this range it is conceivable that the information necessary for encoding a single protein and a few control signals would occupy the entire array of nucleotides. In those chromomeres which are more than 20 times larger, however, any one of several conditions may apply. It is important to discover whether all of the information in a chromomere is necessary for the proper function of the gene. An approach to some aspects of this problem is through the cytogenetic analysis of intragenic duplications and deletions and other chromosomal rearrangements which may or may not interrupt gene function. A group of small rearrangements involving portions of the *white* locus have been used for such a study. Results indicate that the *white* locus normally occupies not band 3C2 as generally believed but a small band between 3C1 and 3C2 of Bridges's map. The thin band is usually very difficult to distinguish but duplication of a portion of the *white* locus allows easier visualization of it. This interpretation makes unnecessary the postulation that the *white* locus occupies only a small fraction of band 3C2. Genetic analysis of the *white* locus rearrangements suggest that the proximal portion of the locus has a regulatory function while the structural gene sequence occupies the distal part of the unit. Some of the small duplications have rather peculiar genetic behavior in that they regularly generate deletions of the *white* locus and sometimes adjacent loci. Occasionally these small segments deleted from the X chromosome are incorporated into other chromosomes of the nucleus. One such insertion into the third chromosome is of particular interest because the control of *white* locus action is entirely changed, apparently by virtue of its new position. The change is not due to variegated position-effect. The genetic instability of the locus and its newly acquired mode of action reflect some important facets of the organization of chromosomes of *Drosophila*.

Cytogenetic studies of the organization of *Drosophila* chromosomes give strong support for the concept that each chromomere of the polytene chromosomes represents, as a general rule, a single module of genetic function (Judd et al., 1972; Hochman, 1971, 1974). This information coupled with the fact that each chromomere of

average size contains about 25,000 nucleotide base pairs creates a paradox central to the understanding of the nature of the genetic program and how it operates in this eukaryote. That a single module of function contains such a large amount of DNA could reflect some degree of redundancy of information coded in a chromomere, though evidence available through the work of Peacock et al. (1974) and Schachat and Hogness (1974) indicates that this is not sufficient to resolve the puzzle. An alternative explanation is that each chromomere contains information for several functions and that the utilization of this information is carried out in such a way that mutations in each of the subdivisions usually behave as a part of a single complementation group. Judd and Young (1974), noting that message sequences in eukaryote cells are apparently derived from larger nuclear RNA precursors, have suggested that the chromomere contains both regulatory and structural gene sequences which can act as a single functional unit primarily because the information is released as a single large transcript. Additionally, since much of this RNA turns over in the nucleus, they postulated that some of the regulatory activity may be carried out by RNA sequences released during the processing of the large messenger precursor transcripts. These regulatory steps would be expected to involve pleiotropic and complementing functions. A more detailed examination of the relationship between the chromomere and the gene in *Drosophila* might help in resolving the puzzle by supplying answers to such questions as:

- 1) Does the gene occupy an entire chromomere (and an adjacent interchromomeric interval) or only some fraction of it?
- 2) Do regulatory mutants possess characteristics that allow them to be recognized as a class distinct from structural gene mutations and if so where do they map in the complementation and recombination maps?
- 3) Do rearrangements which change the position of a gene sometimes result in upsetting the regulation of its activity, and if so what can be inferred as to the nature of the control elements and their functions?

Observations on a group of mutations and chromosome rearrangements involving the *white* locus and its environs provide some partial answers to these and some related questions.

EXPERIMENTAL PROCEDURES AND RESULTS

Chromomeres and complementation groups

Considering first the question of whether the gene and the chromomere are coextensive, we have continued to examine the 3A-3C

region of the X chromosome to discover whether the one-to-one relationship between complementation groups and chromomeres is indeed valid. Recent discoveries have added new complementation groups to those previously known in the 3A-3C region. As reported by Judd and Young (1974) two complementing groups of mutations which cause female sterility have been mapped extremely close to each other in the 3B region of the chromosome. They complement with all of the known lethal groups which map nearby. It is not yet determined precisely what their functional relationship is to each other and to the closely linked lethal loci flanking them. These two complementation groups are good candidates for an example of several functions localized to a single chromomere. It is also possible that they represent complementing alleles of one of the lethal loci, particularly $\ell(1)zw5$ in this region. Additional tests are necessary to determine which of these alternatives is true. This can only be determined by fine structure mapping of the mutants or by identifying and analyzing the gene product or products. Preliminary mapping results show extremely tight linkage between complementing mutants with map distances so small that they are comparable to those ordinarily found between alleles.

Another lethal complementation group, $\ell(1)zw13$, has been added to the 3A region through the experiments of Lim and Snyder (1974). Their intensive screening of mutations induced by several chemicals has produced three representatives of a locus which by our analysis is located between $\ell(1)zw10$ and $\ell(1)zw2$. Recombination tests with these flanking loci show that $\ell(1)zw13$ lies 0.073 map units to the right of $zw10$ and 0.108 map units to the left of $zw2$. This is a region where the correspondence between the chromomeres and the complementation groups was still open to question. Bridges (1938) identified band 3A10 in this section which we had not been able to resolve with the light microscope before publishing our results in 1972. Electron microscope pictures of this region by M. Sorsa and V. Sorsa (Beermann, 1972) showed clearly that band 3A10 exists and we have since been able to resolve it clearly in well stretched polytene chromosome preparations. Chromosome rearrangements which separate 3A9 from 3A10 are not yet available but one which separates 3A8 and 3A9 and one broken between 3A10 and 3B1 give results compatible with placement of the new $\ell(1)zw13$ locus in band 3A9 (Figure 1). The discovery of this locus in region 3B may actually strengthen the one-function : one-chromomere concept because it now reconciles our map with that of Bridges and it appears that each of the chromomeres between the *zeste* and *white* loci contains a genetic unit capable of mutating to a lethal or semi-lethal state. An important point is that efforts to uncover other functions in the 3A-3C segment have been successful and the picture which

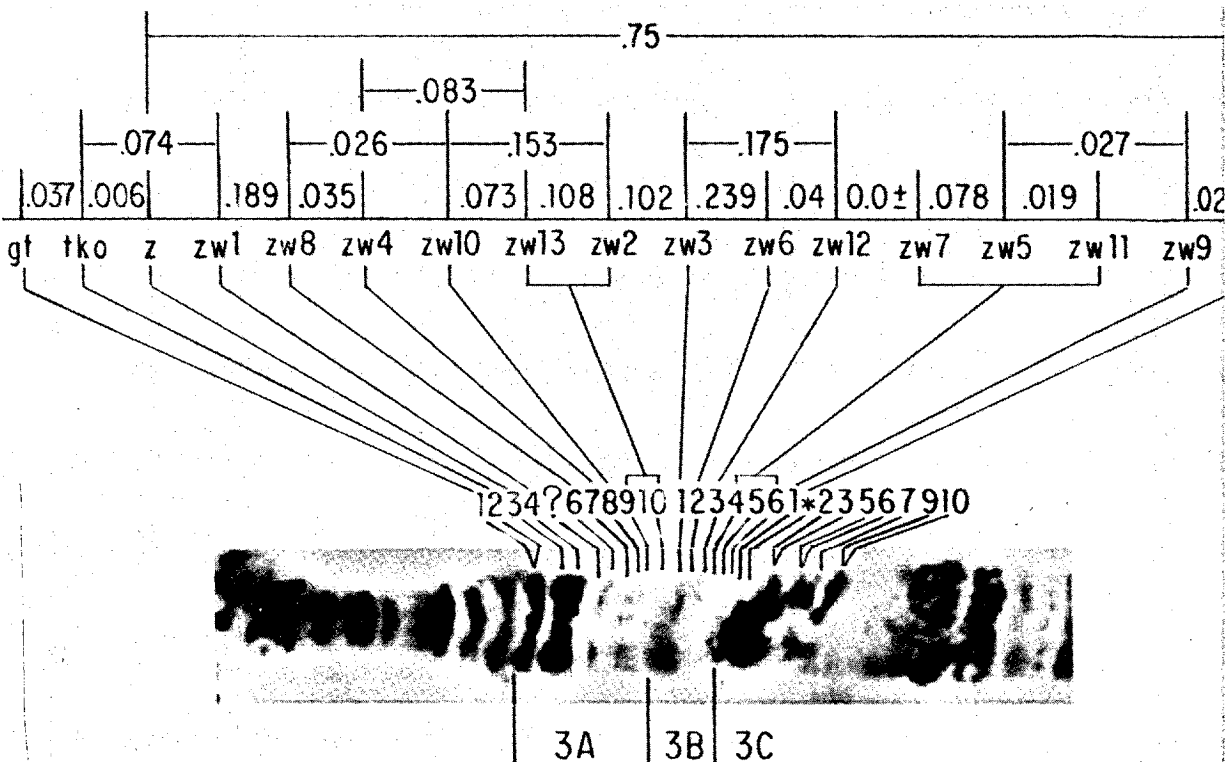


Figure 1. A cytogenetic map of the 3A-3C region of the X chromosome of *Drosophila melanogaster*. The polytene chromosome pictured at the bottom of the figure is from a female of genotype *sc z w^{zm}*. Each of the bands from 3A1 to 3C10 are indicated and correlated where possible with the genetic complementation map of the region. Map distances determined by recombination between complementation units are shown at the top of the figure.

emerges is that these mutants are curiously clustered and extremely tightly linked to a lethal locus.

This cytogenetic approach to determining the number of complementation groups per chromomere, though direct, still suffers from several drawbacks. First, the precise limits of the extent of a chromomere are difficult to determine. Second, there may be entire groups of functional units which do not give mutants that are recognizable by our screening technique (O'Brien, 1973). Though it seems unlikely that a sizeable proportion of the genetic units are incapable of mutating to a lethal or semi-lethal state or to one which causes easily detected morphological changes, it is nonetheless worthwhile to approach the question of the number of functional units per chromomere in a slightly different way.

White locus anatomy and function

An opportunity to do this is presented by the *white* locus mutants and the products of some unusual but regularly occurring unequal crossover events that take place within the *white* locus. These asymmetrical exchanges result in reciprocal duplication (w^{rdp}) and deletion (w^{rdf}) chromosomes. The analysis of a pair of such chromosomes, which were known to be reciprocal because they were recovered in attached-X chromosomes (Judd, 1961), has clearly demonstrated that one chromosome contains a tandem duplication for only the proximal portion of the *white* locus while the other chromosome is missing the same region (Judd, 1964, 1965). It is postulated that the portion of the locus which is involved in the duplication and deletion is a segment which contains regulatory information adjacent to the structural gene sequence. The evidence in support of this idea is gathered from a comparison of phenotypes of various *white* locus mutations and intralocus rearrangements. A map of the fine structure of the *white* locus shown in Figure 2 serves as a reference base for this comparison.

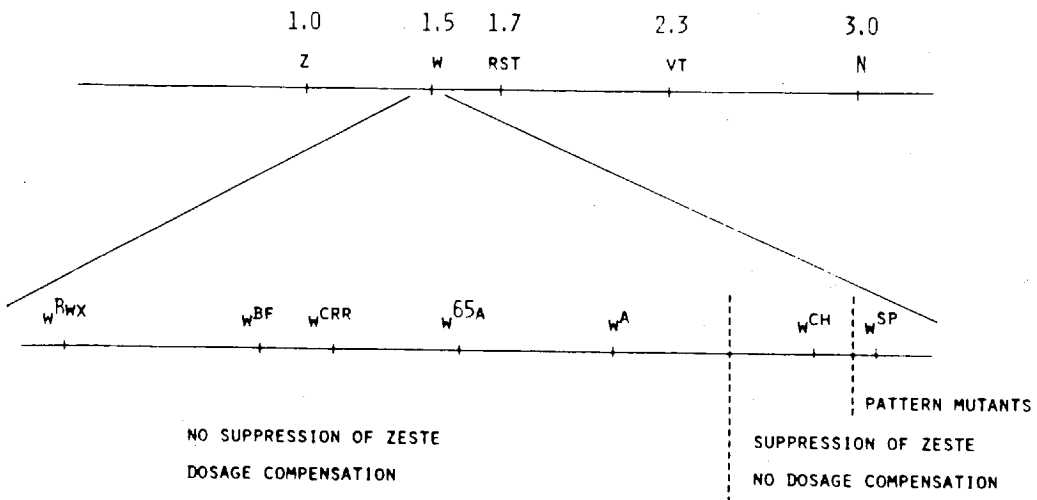


Figure 2. A genetic fine structure map of the *white* locus of *Drosophila melanogaster*. Seven recombination sites are shown (see LeFever, 1973 for details) and functional subgroupings of the mutants are indicated. The relative position of the *white* locus in the distal portion of the X chromosome is shown above.

The model I wish to consider for the arrangement of mutant sites in the *white* locus is that they are distributed in the sequence that specifies a polypeptide structure and in those sequences which play a role in the regulation and co-ordination of the locus. There are

several observations that indicate that mutants located in the right-most or proximal portion of the locus interrupt steps in the regulation of the gene. First, the proximal portion of the locus containing the sites for mutants such as w^{ch} (cherry) and w^{sp} (spotted) is the segment which interacts with the closely linked *zeste* locus, z (Gans, 1953). Mutants located in this proximal section of the gene or a deletion of this section result in a dominant suppression of the mutant *zeste* phenotype (Green, 1959), thus females of genotype $z w^+ / z w^{ch}$ exhibit essentially wild-type eye color, compared to the lemon yellow phenotype of $z w^+ / z w^+$ females.

The duplication of the right-most section of the locus (w^{rdp}) has an effect opposite that of the deletion. It causes a dominant enhancement of *zeste* such that $z w^+ / z^+ w^{rdp}$ females show a *mottled-zeste* phenotype. In males, because the *zeste* phenotype is expressed only when two w^+ genes are present, the $z w^+ / Y$ genotype is wild-type while $z w^{rdp} / Y$ is *zeste*. It is important to note that w^{rdp} is indistinguishable from w^+ in phenotype unless coupled with the z mutation.

It is improbable that a duplication for just a portion of a structural gene sequence would code for a polypeptide chain that has normal function. It follows then that the proximal portion of the locus is not likely to be a part of the sequence which codes for the *white* locus polypeptide. Mutations which map in this section of the gene nonetheless fail to complement with mutants mapping in the rest of the locus, indicating that both structural and regulatory sequences form a single complementation group or cistron.

Second, the *white* locus exhibits dosage compensation so that the male with one w^+ gene has the same phenotype as a female with two. Mutations in the left-most part of the locus retain this compensation ability but those in the proximal portion lose it (Green, 1959). It is of course possible that the ability of a locus to compensate for gene dosage is a property of the structural gene sequence but it appears most logical to assume that this control resides in a regulatory component.

A third observation is that the right-most mutant site mapped in the locus is occupied by the *white-spotted* mutations (Lewis, 1956). Four different mutants are known in this series which change dramatically the distribution of the eye pigment. All four cause pigment to be deposited in small groups of facets scattered across the eyes. The quantity of the pigment is markedly decreased and that which is present seems to be rather yellowish-brown in color compared to the brick-red of the wild-type. The spotting-pattern is completely recessive, in that w^{sp} heterozygous with other *white* alleles produces a uniform brownish eye color; $w^{sp} / w^{deletion}$ heterozygotes,

on the other hand, show a spotted phenotype. Though it is possible that a mutation which changes the gene product could result in these differences in pigment production and deposition, it is more logical to assume that this is due to some change in the control mechanism of the locus.

A fourth set of observations which offer some clues about the regulation of the *white* locus come from combining the intralocus duplications with various *white* mutations in cis arrangement. For example, the duplicated segment of the locus can be shown to contain the site which is changed in *white-eosin* (w^e) as a tandem repeat (w^{e+} , w^{e+}). Green (1963, 1965) showed that it is possible, by crossing over, to insert a w^e mutation into one or the other of these duplicated segments. Interestingly w^{e+} , w^e is wild-type in phenotype while w^e , w^{e+} has an eosin eye color. The same is probably true for w^{ch} where it is known that w^{ch+} , w^{ch} is wild-type but w^{ch} , w^{ch+} is apparently indistinguishable from w^{ch} (Judd, 1965). The supposition is that only when the non-mutant control element segment of the duplication is adjacent to the structural sequence to its left is normal function of the locus allowed. Such a polar effect is similar to that seen in mutants affecting the operator and promoter segments of a prokaryote operon. Under ordinary circumstances such polar effects are not characteristic of mutations in eukaryotes and it is probably exhibited in these cases only because of the tandem duplication for a part of the locus.

The cytogenetic size of the *white* locus

Another interesting *white* locus mutation that exhibits all of the characteristics of an intralocus duplication is w^{zm} . This mutant originated as a single scute mottled-eyed male from a female of genotype $sc\ z\ w^+ ec\ ct / w^{bf}$ (sc = scutellar bristles absent; z = zeste eye; w^{bf} = white-buff eye; ec = echinus eye; ct = cut wing). There is evidence that the mutation is the result of an unequal crossover in the *white* locus which produced a duplication for the proximal part of the locus (Judd, 1963) very similar to w^{rdp} described above. As in the case of w^{rdp} , w^{zm} maps at the right of w^a and acts as a dominant enhancer of z . Unless coupled with the mutant z , however, it produces a wild-type eye color.

The w^{zm} chromosome offers another cytogenetic approach to the question of what part of a chromomere is occupied by a gene. Unlike w^{rdp} which cytologically appears to add little material to the chromomere in which it is presumed to be located, w^{zm} does cause a visible addition to the already considerable amount of DNA in the distal part of section 3C. Since it can be conclusively demonstrated that w^{zm} is duplicated for only a part of the *white* locus, it should be possible to estimate how much of the chromomere

is occupied by the *white* locus by determining how much material appears to be added cytologically by the w^{2m} duplication.

The first step in this analysis is determining as accurately as possible what the standard cytological picture for this region of the chromosome is. Bridges (1938) pictured 3C1 as a rather prominent band of above average thickness followed by a very heavy doublet, bands 3C2,3. Cytogenetic analysis carried out by several investigators over the years have placed the *white* locus in any or even all of bands 3C1,2 and 3. The most thorough and best documented study is that of Lefevre and Wilkins (1966) which indicated that *white* resides in band 3C2. The picture becomes somewhat clouded again, however, when it is noted that in some instances a very faint but distinct band can be seen between 3C1 and the 3C2,3 doublet. Lefevre and Green (1972) published a photograph of a well stretched $y w spl$ chromosome showing such a band (Figure 3). They state that this is a unique observation but they left open the possibility that some or all of the w locus occupies a position between 3C1 and 3C2. However, since an intra-locus deficiency derived as an unequal crossover appears cytologically to have a diminished 3C2,3 doublet, they favored 3C2 as the position of w . Sorsa et al. (1973) noted a somewhat thicker band following 3C1 in a chromosome that has an inversion at or near the *white* locus that transfers the 3C2,3 doublet to section 12B. Their genetic tests showed that the w^+ function was not associated with the inverted segment and they then considered the possibility that the inversion break passed through band 3C2 leaving a small part of it in the original position and transferring the remainder to 12B. The obvious conclusion, if w resides in band 3C2 and if 3C2 has been fractured, is that the *white* gene can occupy only a small part of the material comprising that band.

My studies on the rearrangements with break points in and about the *white* locus lead me to a different interpretation. It is clear from examining w^{2m} and derivatives of it, that the increase in material in this region is by thickening of a band normally positioned between 3C1 and 3C2 but which is extremely difficult to separate from 3C2 in wild-type chromosomes. It is relatively easy to visualize this band in some stocks, particularly the intralocus duplication stocks, w^{2m} , w^{2l} , and w^2 where the band is easily visible, being rather thick and quite prominent in stretched preparations (Figure 3). It would appear that w might occupy this band, which I shall call 3C1.5. If the *white* locus is indeed located in this thin band, the evidence indicates that it occupies a major portion of it, since a duplication which genetically includes only a part of the locus increases considerably the width of the band. This evidence is circumstantial and other interpretations are

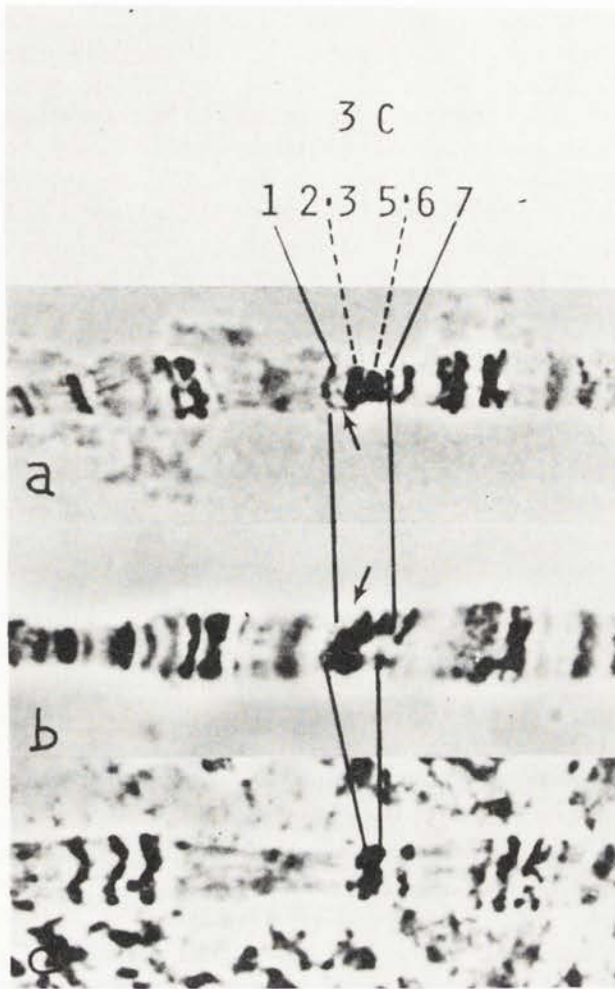


Figure 3. Polytene chromosomes of *Drosophila melanogaster* showing: (a) bands in section 3C in a well stretched preparation of an unsynapsed *y w spl* chromosome from a *y w spl/Tp(1)L2* female. The arrow indicates a thin band clearly visible between bands 3C1 and 3C2. This photograph was supplied by Dr George Lefevre Jr. (b) section 3C in a female of genotype *sc z w^{zm}*. *w^{zm}* is a duplication for the proximal portion of the *white* locus. The arrow indicates a rather thick band between 3C1 and 3C2. (c) section 3C in a female homozygous for *Df(a)w^{zh}* which was derived from *w^{zm}*. Note that the heavy doublets 3C2, 3 and 3C5, 6 are deleted.

possible but it is interesting to note that the rearrangement used by Sorsa et al. (1973) was derived from a chromosome which contained an intralocus triplication very similar in origin and genetic makeup to the duplications *white* described above. If my interpretation is correct, the *white* locus would continue to map in the non-inverted

segment of this chromosome even though the 3C2,3 doublet was clearly moved to a new position. This opens still another question about the functions associated with the 3C2,3 and 3C5,6 doublets. If the *white* locus is not located in 3C2, there is no functional group yet identified for that chromomere and this leaves only two known loci, *roughest* (*rst*) and *verticals* (*vt*) occupying this region made up of four bands. Lefevre and Green (1972) and Kaufman (1971) present evidence that the 3C2,3 and 3C5,6 doublets are tandem duplications of each other which may explain the apparent paucity of functional units detected there.

Gene superstructure and gene control

Turning now to the question of whether rearrangements of the position of a gene affect the way its function is regulated, an analysis of some of the *white* locus changes is also helpful. I do not wish to consider the phenomena of variegated position effect but rather I will deal with one of the mutable systems that involves changes in *white* locus position and function.

A more detailed description of the origin and behavior of the mutable system will appear elsewhere (Judd and Lundelius, in preparation) and so I shall deal only with those aspects which are pertinent to the question at hand, namely changes in the regulation of gene function due to changes in gene position. It is important, however, to outline in a general way the origin and behavior of one of the *white* locus transpositions.

The transposition, designated $Tp(1 \rightarrow 3)w^{zh}$, and which will be referred to simply as w^{zh} was derived from a w^{zm} chromosome through several steps. The w^{zm} change, as described above, is most likely a duplication for a section of the *white* locus located to the right of w^a (Figure 2). It mutated to a form which causes absence of all pigment from the eyes. This *white* type (designated w^z) proved upon genetic analysis to be a double mutant at the *white* locus. It continues to have the duplication which typifies w^{zm} and in addition there is a mutation which is at or very near the site of w^a . The w^z chromosome is highly mutable, causing a variety of changes at the *white* locus. These changes generally are transmitted through the germ line as clusters of identical changes, indicating that they arise as a premeiotic event. One of the classes of changes which are recognized, are deletions of the *white* locus and in some cases adjacent loci as well. These are easy to recognize because the removal of the *white* locus, specifically the rightmost portion of it, causes a suppression of the *zeste* phenotype. In a series of experiments designed to recover and analyze such deletion events, a male was discovered which had apparently reverted from $sc\ z\ w^z$ back to the w^{zm} phenotype. It was from a line

established from this male that a *sc* ' w^+ ' *rst* male appeared. Analysis of this apparent reversion of the *white* locus to wild-type with the simultaneous mutation of the nearby *roughest* locus showed that the X chromosome had been deleted for bands 3C2 through 3C6 inclusive (Figure 3) and that at least the portion carrying w^+ had been inserted into the left arm of chromosome 3. The deletion in the X chromosome is lethal in males which do not also carry the transposed segment. Males with the deleted X chromosome plus the transposition in one chromosome 3 show a completely wild-type phenotype for *white* but are mutant for *roughest* and also exhibit a mild *verticals* effect. A truly surprising phenotype is found in such males when the chromosome 3 transposition is made homozygous. The eye color is unlike any previously reported, being wild-type brick red in the facets of the periphery of the eye and grading to *zeste* at the center of the eye. *Roughest* is still expressed in an extreme form but the fly is wild-type for the *verticals* phenotype.

Focusing first on the eye phenotype, the curious pattern of pigment distribution is most astonishing. It is almost as if the w^+ gene were not functioning in the eye itself but is forming its product in surrounding tissues with diffusion of the product into the eye. We know that the transposed *white* gene is capable of producing a normal product at the proper time and in the proper tissue when it is present in a single dose because the phenotype of the eye in flies heterozygous for the transposed locus is indistinguishable from wild-type. A possible clue to this dosage effect comes with the observation that the halo phenotype is produced by the homozygous transposition only when the X chromosome carries mutant alleles of both the *z* and *w* loci. It should be recalled that the expression of the *zeste* phenotype is directly dependent on the number of doses of w^+ . This relationship is carried further with the analysis of one of the variants of the transposed element. A dominant form of w^{zh} appeared which causes the halo effect when present in only one dose. When the dominant is made homozygous (w^{zhD} / w^{zhD}), the halo phenotype is produced when the X chromosome is mutant at either the *z* or *w* locus.

One of the characteristics of the transposed segment is its continuing mutability. There are two aspects of the mutability which appear to be circumstantially related. First, the transposition has the quality of moving to other positions in the genome and second, the way in which the genes in the transposed segment function may be modified on a semi-permanent basis such that the change is inherited and a line with the new characteristics can generally be established and maintained by selection. Changes of this latter type with regard to the *white* locus include variations in the type and distribution of pigment in the eyes, some of which

are quite astonishing. One example is a strong reduction in the pigmentation of the eye to the shade of light yellow with the removal of the halo pattern at the periphery of the eye except for a small section in the posterior ventral quarter which always contains a streak of wild-type pigment. In some variants the halo pattern is retained but the pigment is markedly reduced to give an orange-yellow halo effect. In others, the pigment distribution becomes more uniform or is distributed in a mottled pattern throughout the eye.

During the early part of the analysis of the transposition it was believed that only the *white* locus and possibly *verticals* had been inserted into the third chromosome because the *roughest* phenotype in the deficiency males persisted when the transposition was present. During the analysis, however, a line appeared which was *rst⁺* and represented the derepression of the *roughest* locus in the transposed element. I assume that *rst⁺* had been present in the transposition all along but because of its new position in chromosome 3 it failed to receive or respond to the regulatory signals as it ordinarily would have. In this respect both the *white* and *roughest* genes are upset in some step in the regulatory network because of their new position. The fact that under some circumstances or in some variant forms the normal gene product is apparently formed at the proper place and time indicates that the structural sequences of these genes have not been impaired. The aspect of the regulatory elements which have been modified by the new gene position are only semi-permanent. The change which has been imposed is inherited through the germ cells much like an ordinary mutation but it is clearly not something as stable as a change in nucleotide sequence. The change appears to be much like that imposed on genes during the process of cell differentiation. It could involve changes in the superstructure of DNA in the fashion postulated by Cook (1973) who suggested that cellular differentiation is established by the development in chromosomes of a particular superstructure configuration. The mechanism operating in the system described here enables these changes to be replicated and transmitted through the germ cells as well as in somatic tissues. Let me emphasize a point made earlier that the changes described in this mutable system are not due to variegated position effects where genes located in euchromatin become transported to the vicinity of a block of heterochromatin. This system has many of the characteristics of the mutable genes in *Drosophila* described by Green (1967, 1973) and in maize described by McClintock (1956).

DISCUSSION

I have presented data which I believe support the idea that the module of genetic function in *Drosophila* is rather large and complex; it is considerably more than the DNA sequence which specifies a polypeptide chain. A logical interpretation of these data is that some of the information in a chromomere is important to the way a gene functions and to the way it interacts with other genes even though that information does not concern the specification of the structure of the polypeptide product of the gene. This information could of course be that which is used either in transcription or translation to regulate the quantity or distribution of the gene product. I have tried to develop the possibility, however, that there are other avenues for communication between genes in addition to the protein-nucleic acid interactions that are typical of gene repression and derepression in prokaryotic systems.

In those eukaryotes where the size of the initial transcript has been measured, it can be of the order of 20,000-50,000 nucleotides long, while the portion which remains after processing in the nucleus and which is transported to the cytoplasm as mRNA is very much smaller, say 500-4,000 nucleotides (Molloy et al., 1974). This opens the possibility that some of the information in the chromomere is transcribed and used post-transcriptionally in one of two ways. Some of the information may be necessary for the processing of the transcript itself and any mutations in this segment might be expected to act in a cis-dominant way because a functional mRNA might not be produced even though the message sequence is transcribed and contains the information for a proper polypeptide. A second group of RNA sequences may function when they are released during the processing of the mRNA to repress or derepress other genes of the nucleus. In this latter class, a mutation might be expected to act pleiotropically if it should fail to carry the proper signal to another gene or genes and at the same time would interfere with the maturation of the mRNA in its own transcript. The interesting prediction can be made that such mutants may be identified as lesions in a regulator locus whose action is limited to the nucleus in which it resides and thus in a heterokaryon carrying normal and mutant nuclei the mutant effect would be autonomous for each nucleus even though the mutant is recessive. An example of a mutation in a gene which controls only the synthesis of enzymes coded in the nucleus in which it is located is *scn^c* (sulfur-control constitutive) in *Neurospora* (Burton and Metzenberg 1972; Dietrich and Metzenberg, 1973). This is an excellent candidate for a mutation in a sequence which is transcribed but not translated.

I have also suggested that some of the information in a chromomere need not be transcribed to function in a regulatory role but rather may be responsible for conferring structural modifications on the DNA of the entire chromomere which cause the gene to be repressed or derepressed or to respond only to limited types of repressors or derepressors in particular cell types depending upon such things as the developmental history of the cell. It appears that to some degree this superstructural modification, if such be the case, can be influenced markedly when genes are moved to new positions in the genome. This could be in part due to these changes extending beyond the boundaries of a single chromomere. This might indicate that there is a supragenetic organization that ties blocks of genes together under a rather general type of regulation much as in the X chromosome inactivation of mammalian females. I consider it highly significant that the changes which upset the regulation of genes placed in new positions behave not as mutations in the base sequence of the DNA but rather as semi-permanent modifications of the way a gene responds to its environment. These are changes much like those conferred during cell determination and differentiation. The unusual nature of some of the examples I have cited is that these changes can be and most frequently are inherited through the gametes rather than being limited to somatic tissues as are the usual changes imposed during development.

ACKNOWLEDGMENTS

The excellent assistance of Margaret W. Shen is gratefully acknowledged. I wish to thank Dr George Lefevre Jr for supplying one of the photographs in Figure 3. Appreciation of the help of M.W. Young and T.K. Johnson in critical reading of the manuscript is acknowledged. This work was supported by National Science Foundation Grant GB-34887 and by U.S. Public Health Service Grant GM-00337.

REFERENCES

- Beermann, W. 1972. Chromomeres and Genes. *In* Developmental Studies on Giant Chromosomes. Results and problems in cell differentiation (W. Beermann, ed.) Vol.4, p.1. Springer-Verlag, Berlin.
- Bridges, C.B. 1938. A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. *J. Heredity* 29: 11.
- Burton, E.G. and R.L. Metzenberg. 1972. Novel mutation causing derepression of several enzymes of sulfur metabolism in *Neurospora crassa*. *J. Bacteriol.* 109: 140.
- Cook, P.R. 1973. Hypothesis on differentiation and the inheritance of gene superstructure. *Nature* 245: 23.

- Dietrich, P.S. and R.L. Metzberg. 1973. Metabolic suppressors of a regulatory mutant in *Neurospora*. *Biochem. Genet.* 8: 73.
- Gans, M. 1953. Etude genetique et physiologique du mutant z de *Drosophila melanogaster*. *Bull. Biol. France Belg. (Suppl.)* 38: 1.
- Green, M.M. 1959. Spatial and functional properties of pseudo-alleles at the *white* locus in *Drosophila melanogaster*. *Heredity* 13: 302.
- . 1963. Unequal crossing over and the genetical organization of the *white* locus of *Drosophila melanogaster*. *Z. Vererb.* 94: 200.
- . 1965. Genetic fine structure in *Drosophila*. In *Genetics Today* (S.J. Geerts, ed.), Vol.2, p.37. Pergamon Press, Oxford.
- . 1967. The genetics of a mutable gene at the *white* locus of *Drosophila melanogaster*. *Genetics* 56: 467.
- . 1973. Some observations and comments on mutable and mutator genes in *Drosophila*. *Genetics* 73, Suppl.: 187.
- Hochman, B. 1971. Analysis of chromosome 4 in *Drosophila melanogaster*. II: Ethyl methanesulfonate induced lethals. *Genetics* 67: 235.
- . 1974. Analysis of a whole chromosome in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 581.
- Judd, B.H. 1961. Formation of duplication-deficiency products by asymmetrical exchange within a complex locus of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 47: 545.
- . 1963. The genetic fine structure of the mutants z^m and z^l in *Drosophila melanogaster*. In *Genetics Today* (S.J. Geerts, ed.), Vol.1, p.3. Pergamon Press, Oxford.
- . 1964. The structure of intralocus duplication and deficiency chromosomes produced by recombination in *Drosophila melanogaster*, with evidence for polarized pairing. *Genetics* 49: 253.
- . 1965. Chromosome pairing and recombination in *Drosophila melanogaster*. *Genetics* 52: 1229.
- Judd, B.H., M.W. Shen and T.C. Kaufman. 1972. The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics* 71: 139.
- Judd, B.H. and M.W. Young. 1974. An examination of the one-cistron: one-chromomere concept. *Cold Spring Harbor Symp. Quant. Biol.* 38: 573.
- Kaufman, T.C. 1971. The genetic anatomy and function of the 3A1-3C7 region of the X chromosome of *Drosophila melanogaster*. Dissertation. The University of Texas at Austin.
- LeFever, H.M. 1973. Analysis of three *white* mutants resulting in two new recombination sites at the *white* locus in *D. melanogaster*. *Drosoph. Inform. Serv.* 50: 109.

- Lefevre, G.Jr and M.O. Wilkins. 1966. Cytogenetic studies on the white locus in *Drosophila melanogaster*. *Genetics* 53: 175.
- Lefevre, G. Jr and M.M. Green. 1972. Genetic duplication in the white-split interval of the X chromosome in *Drosophila melanogaster*. *Chromosoma* 36: 391.
- Lewis, E.B. 1956. An unstable gene in *Drosophila melanogaster*. *Genetics* 41: 651.
- Lim, J.K. and L.A. Snyder. 1974. Cytogenetic and complementation analyses of recessive lethal mutations induced in the X chromosome of *Drosophila* by three alkylating agents. *Genet. Res.* In press.
- McClintock, B. 1956. Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol.* 21: 197.
- Molloy, G.R., W. Jelinek, M. Salditt and J.E. Darnell. 1974. Arrangement of specific oligonucleotides within poly(A) terminated HnRNA molecules. *Cell* 1: 43.
- O'Brien, S.J. 1973. On estimating functional gene number in eukaryotes. *Nature New Biol.* 242: 52.
- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 405.
- Schachat, F.H. and D.S. Hogness. 1974. Repetitive sequences in isolated Thomas circles from *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 371.
- Sorsa, V., M.M. Green and W. Beermann. 1973. Cytogenetic fine structure and chromosomal localization of the white gene in *Drosophila melanogaster*. *Nature New Biol.* 245: 34.

The structure and regulation of eukaryotic genomes approached by studying HnRNA

James E. Darnell Jr

A case is presented for the necessity to approach a study of eukaryotic chromosomal organization with the lessons of bacterial genetics firmly in mind, but with the expectation that biochemical rather than genetic experiments are more likely to give information on distribution of sequences encoding mRNA and potential regulatory functions. To this end studies on the organization of HnRNA, the giant nuclear RNA molecules which are in part thought to be mRNA precursors have been undertaken.

After the recognition in the early part of this century of the central role of chromosomes in heredity, a large part of the biologic community has understandably remained engaged in studying the composition and the detailed cytological and genetic structure of chromosomes. An important direction of the present work on chromosomes, which has been imparted by modern biologists, (particularly of the Jacob-Monod school), concerns the molecular details of the *regulation* of structural gene expression. The biochemistry of mRNA synthesis and fate are central subjects in such molecular studies.

Since the early 1960s considerable progress has been made in studying the pathways of RNA biosynthesis in eukaryotic cells by examining newly made (rapidly labeled) RNA extracted from growing cultured cells (for review, see Darnell, 1968, 1974). First came the discovery that rRNA was made from a high molecular weight nucleolar precursor (Scherrer et al., 1963; Perry, 1962; Penman et al., 1966, and Maden, 1971). It was then found that tRNA was also processed from a higher molecular weight RNA species termed pre-tRNA (Burdon and Clason, 1969; Bernhardt and Darnell, 1969).

The recalcitrant member of the cell's family of major ribopolynucleotides in joining the list of post-transcriptionally modified molecules was mRNA. Although a very high molecular weight 'DNA-like' (U substituted for T) nuclear RNA fraction (Sibatani et al., 1962; Scherrer et al., 1963; Georgiev and Mantieva, 1962; Ellem and Sheridan, 1964; Yosikawa-Fukada et al.,

1965; Houssais and Attardi, 1966; Soeiro et al., 1966) was proposed over 10 years ago as a candidate for precursor to the DNA-like mRNA (Scherrer et al., 1963; Penman et al., 1963; Scherrer and Marcaud, 1965), no convincing evidence was forthcoming that such a relationship was actually true.

THE DERIVATION OF mRNA FROM HnRNA

Recent work has identified a number of sequences that are shared between HnRNA and mRNA, substantially increasing the likelihood that mRNA is derived by post-transcriptional modification of the higher molecular weight nuclear RNA. Two general types of shared sequences have been identified in mRNA and HnRNA:

(1) Sequences that are transcribed from regions of the cell genome (structural gene region) that encode polypeptides and (2) a post-transcriptionally added segment. The first possible structural gene sequences found in both HnRNA and mRNA were virus-specific sequences transcribed from the integrated DNA of tumor viruses (Benjamin, 1966; Sambrook et al., 1968; Lindberg and Darnell, 1970; Tonegawa et al., 1970; Wall et al., 1973). More recently, sequences present in hemoglobin and immunoglobulin (Imaizumi et al., 1973; Williamson et al., 1973; Melli and Pemberton, 1972) have been found in HnRNA. The present evidence in these situations does not prove that the mRNA sequences actually are first transcribed as the longer HnRNA molecules, but are consistent with this hypothesis.

After HnRNA has been transcribed there is *post-transcriptional* addition of a homopolymeric segment (about 200 nucleotides long) of poly(A) (Edmonds and Abrams, 1959, 1962; Edmonds et al., 1971; Mendecki et al., 1972; Darnell et al., 1971a; Jelinek et al., 1973) by non-DNA dependent extension (Philipson et al., 1971; Birnboim et al., 1973) of the 3' terminus of at least a portion of the HnRNA molecules. This unusual homopolymeric sequence (Molloy and Darnell, 1973) occurs in the majority of mRNA molecules of all eukaryotic cells (Kates, 1970; Lim and Canellakis, 1970; Edmonds et al., 1971; Lee et al., 1971; Darnell et al., 1971b; Firtel et al., 1972). Furthermore an adenosine analogue, 3'dA (deoxyadenosine or 'cordycepin') prevents the addition of poly(A) (Darnell et al., 1971a; Mendecki et al., 1972) and the occurrence of some necessary post-transcriptional event(s) leading to the cytoplasmic appearance of mRNA (Penman et al., 1970; Darnell et al., 1971a; Philipson et al., 1971; Adesnik et al., 1972). Finally it has recently been shown possible that most poly(A) added to HeLa cell HnRNA appears in the cytoplasm (Jelinek et al., 1973; Puckett, Chambers and Darnell, unpublished). It appears likely therefore

that poly(A) addition is a necessary step in mRNA formation for those mRNA molecules containing the homopolymeric segment.

The experiments which show shared and/or conserved sequences between mRNA and HnRNA make it likely that most mRNA is, in fact, not synthesized independently, but is first formed as HnRNA and post-transcriptional events then appear to be involved in the majority of mRNA biogenesis. It is also possible, however, that a minority of mRNA molecules, including histone mRNA which lacks poly(A) (Adesnik and Darnell, 1972; Greenberg and Perry, 1972) may be transcribed directly or processed in a manner different from the poly(A) pathway.

REGULATION OF mRNA FORMATION

This new information on mRNA biosynthesis is cause for a reassessment of the possible modes of regulation of mRNA formation in eukaryotic cells.

The framework within which the regulation of eukaryotic mRNA formation should be considered naturally derives from bacterial physiology where conclusive evidence about regulation of gene expression exists. A bacterium controls protein synthesis by regulating the *initiation* of mRNA synthesis (Jacob and Monod, 1961; review in *Cold Spring Harbor Symp. Quant. Biol.* 35, 47, 1970; Zubay et al., 1970, Greenblatt and Schleif, 1971). The gene transcripts, i.e. newly-made RNA molecules, function as mRNA *even before* chain completion of the RNA has been achieved (Miller et al., 1970). Thus only minor post-transcriptional modifications of mRNA could possibly be required and even if these occurred they would occur at the 5' end of new RNA molecules, not the 3' end.

mRNA metabolism in eukaryotic cells presents, then, two major mechanical differences from bacteria. First, there are extensive post-transcriptional, 3' terminal modifications, and, second, at least some of the mRNA generally has a long half-life compared to bacterial cells (Penman et al., 1963; Singer and Penman, 1973; Perry and Kelley, 1973). Both transcriptional regulation of the number or kind of HnRNA molecules synthesized as well as post-transcriptional regulation (? nuclear) of the number and kind of mRNA molecules fashioned from HnRNA would appear to be possible (Table 1). In addition, because of the long mRNA half-life, interruption or change in efficiency of functional cytoplasmic mRNA:ribosome complexes may also be required to alter the rate or abruptly stop established protein synthesis (McAuslan, 1963; Tompkins et al., 1969; McCormick and Penman, 1968). This might be termed *translation modulation*, reserving the term *regulation* for those potential regulatory events concerned with mRNA formation.

TABLE 1
Formation and regulation of mRNA

Location	Life history of mRNA	Possible levels of regulation
Nucleus	1. Transcription of HnRNA	1. Transcriptional
Nucleus	2. Poly(A) addition	
Nucleus	3. Cleavage	2. Post-transcriptional
Nuc./cyto.	4. Transport of mRNA	
Cytoplasm	5. Translation into protein	3. Translational modulation

GENETICS OF REGULATORY GENES

In this article we wish to consider the types of experimental approaches in eukaryotic cells that can elucidate the mechanisms that control how many mRNA molecules are formed and delivered to the protein synthesizing apparatus. In bacteria, genetic identification of regulatory mutations was first accomplished (Jacob and Monod, 1961) followed by biochemical verification of the mode of action of each genetic element (see *Cold Spring Harbor Symp. Quant. Biol.* 35 for review) (Table 2). Mutations of regulatory genes were characterized by showing that regulatory protein products from one set of genes could act on another inhibited set of genes (trans) to control the expression of structural gene products. Regulatory genes may produce either negative regulatory proteins (repressors) which prevent specific mRNA transcription (Zubay et al., 1970) or positive regulatory proteins (activators) which increase specific mRNA transcription (Greenblatt and Schleif, 1971). In addition specific regulatory sites have been genetically identified which cannot function unless they are very closely linked (cis) to the structural genes which they regulate. These sites (promoters, operators) have now been shown to be specific regions of the bacterial chromosomal DNA which binds either RNA polymerase or regulatory proteins (Gilbert, 1967 and *Cold Spring Harbor Symp. Quant. Biol.* 35). In the functioning of all these elements in bacteria, it is most important to remember that the regulatory decision concerns the promotion of, or prevention of, transcription of an RNA molecule from a region (operon) of the bacterial

TABLE 2
Possible regulatory elements in eukaryotic cells and results of mutational defects in such elements

Mutation in	Mutant should act		Result of mutation on structural gene product
	cis	trans	
1. Regulatory protein			
DNA binding	Repressor ^a	+	Formed constitutively
	Activator ^a	+	Not formed
HnRNA binding	Negative-acting ^b	+ ^c	Formed constitutively if HnRNA transcribed ^c
	Positive-acting ^b	+ ^d	Not formed ^d
2. Regulatory binding in nucleic acid			
DNA	Operator ^a	+	May be formed constitutively or not formed; further genetic tests needed
	Promoter ^a	+	
HnRNA	For negative-acting protein	+	Formed constitutively if HnRNA is made
	For positive-acting protein	+	Not formed
3. Structural gene			
		+	Not formed

a. Known to exist in bacteria; two model systems are e.g. the β -galactosidase operon (negative controlled; Zubay et al., 1970) and the arabinose operon (positive controlled; Greenblatt and Schleif, 1971).
 b. Negative action on HnRNA would be equivalent to repressor-like activity; positive action equivalent to activator-like activity. Since these proteins might act on many types of HnRNA molecules, truly defective mutations might be lethal or at least pleiotropic.
 c. Structural gene product might be expressed if hypothetical negative-acting protein concentration was too low in heterozygote; even homozygosity for deficient negative-acting protein would allow structural gene expression only if HnRNA were transcribed.
 d. Might appear mutant only if homozygous since in heterozygote some good positive-acting protein might suffice; alternatively, if a positive-acting protein were multimeric, one bad subunit might be enough to inactivate and mutation would appear in heterozygote.

chromosome - this RNA transcript is the mRNA (Miller et al., 1970). In eukaryotic cells we propose (Darnell et al., 1971a, 1973) that the HnRNA transcript contains the mRNA but only after post-transcriptional modification does the transcript become the mRNA. Thus two basic possibilities for regulation of mRNA production exist: (1) The synthesis of HnRNA transcripts containing mRNA might be regulated using the principle established in bacteria of recognition sites in the DNA and regulatory proteins which combine with such sites; (2) It is also possible that in addition to or instead of transcriptional regulation, differential conservation of mRNA from the HnRNA transcripts may occur. This would be post-transcriptional regulation of mRNA production.

Because of the two possibilities in the level of regulation of mRNA synthesis, genetic studies of the type which revealed so much about bacterial regulation may not provide easily interpreted results for eukaryotic cells in general and mammalian cells in particular. For purposes of discussion, assume that in eukaryotic cells positive and negative transcriptional regulatory genes and sites (promoters, operators) on the DNA for interaction with the regulatory proteins exist as they do in bacteria (Table 1). Suppose further that proteins exist which function in mRNA biosynthesis at a post-transcriptional level and that binding sites for these proteins exist on the HnRNA molecules. Such post-transcriptional regulatory proteins might have either a positive (activator-like) or negative (repressor-like) regulatory role in mRNA manufacture.

Table 2 gives an outline of the expected characteristics of mutations in these various hypothetical regulatory elements. To illustrate the difficulties in deciding the molecular basis of mutational defects in the regulatory systems of eukaryotes, let us consider mutations causing defects, for example, in positive regulatory proteins (activators) or the binding sites for such proteins. Defective mutants in either the gene(s) for positive-acting proteins or their binding sites would result in inactivation of the structural genes *whether the locus of action was transcriptional or post-transcriptional*. Demonstration of the intactness of the structural gene through recombination would suggest that such mutations were regulatory, but could not prove the molecular site (DNA or HnRNA) of the regulation nor whether the mutation was in a regulatory protein or in its binding sites. Detailed genetic studies involving cis and trans tests and fine structural mapping (which involves the repositioning of mutant genetic elements) coupled probably with studies on specific HnRNA synthesis will eventually be required to settle such issues. In

mammalian cells, the genetic technology for cis and trans tests might be accomplished by cell fusion techniques but recombination to test for structural gene damage or excision and replacement of hypothetical defective regulatory elements seems simply out of reach for the present. It is possible that genetic identification of regulatory mutants may be accomplished in *Drosophila* at such finely studied loci as *rosy*, the gene for xanthine dehydrogenase (Glassman, 1965; Chovnick, 1966). Regulatory genes can be defined in molds, e.g., *Neurospora* (Marzluff and Metzgenberg, 1968) and here the fine structure genetics can be more easily performed. However, almost no information on the biochemical events of mRNA formation is available in *Neurospora*. Among lower eukaryotic cells the amoebic form of *Dictyostelium* is among the most extensively studied biochemically and the results on mRNA formation by these cells demonstrate possibly a simpler situation than in mammalian cell mRNA manufacture (Firtel et al., 1972). Slime mold cells have poly(A) in mRNA as well as a nuclear RNA fraction but the nuclear RNA is only about 500 bases longer than is the mRNA. Thus if longer length of the HnRNA means more complicated regulation, then the study of lower eukaryotic cells such as *Dictyostelium* or *Neurospora* may provide important guidelines but not necessarily final answers about the complexity of regulation in mammalian cells. Even more than in the study of bacterial physiology it will be necessary to approach mammalian cell regulation *by design* through a study of molecular events in mRNA formation. Only with the development of genetic techniques which will allow fine structure mapping in mammalian cells (e.g., through somatic crossing over, transduction, etc.) does it appear that mammalian cell regulatory genes will receive adequate genetic description. A corollary to the limitations of genetic studies with mammalian cells are the limitations of studies of the anatomy of DNA molecules without reference to their function. For example, a great deal at this conference has been said about sequences in eukaryotic DNA which are repeated (or nearly repeated) many times in the genome of all eukaryotic cells (Britten and Kohne, 1968). A great effort has gone into characterizing and mapping the distribution of such sequences in DNA molecules (Davidson et al., 1973) on the assumption that eventually it will be shown that such regions are in some way related to gene regulation (Britten and Davidson, 1969). Many of these repeated sequences are transcribed into HnRNA (Soeiro and Darnell, 1970; Pagoulatos and Darnell, 1970; Darnell and Balint, 1970; Scherrer et al., 1970; Perry et al., 1970; Shearer and McCarthy, 1967; McCarthy and Duerksen, 1970) and if post-transcriptional regulation occurs, the potential regulatory function of these sequences may occur after transcription and studying DNA anatomy will never reveal that fact.

Even if the repeated sequences (or some of them) are related *only* to transcription (Britten and Davidson, 1969), the localization of these sites with respect to transcriptional initiation or termination signals may be much easier through the study of HnRNA structure, since HnRNA molecules at least reflect a portion of the genome active in transcription.

USE OF HnRNA TO STUDY CHROMOSOME STRUCTURE

Our laboratory has been attempting to gain information about the structure of initial transcripts from mammalian cell chromosomes. Eventually to settle the locus (loci) of regulation in mammalian cells (transcriptional, post-transcriptional or both) what must be answered for as many genetic regions as possible is whether, during increased production of mRNA from a specific gene, there is an increased number of transcripts from the chromosomal regions bearing the gene, or an increased conservation of mRNA from the HnRNA transcripts, or a combination of the two.

To begin the task of identifying and characterizing transcripts which will eventually lead to an mRNA product, *not* already processed products on the way to becoming mRNA, or HnRNA not destined to give rise to mRNA, a separation must be achieved of large nuclear RNA molecules which contain a region corresponding to a structural gene which will be used. A beginning of this task is possible, by selecting the largest nuclear RNA molecules which contain the 200 nucleotide segments of polyadenylic acid at their 3' termini (Molloy et al., 1974). Although conclusive evidence is lacking that every poly(A) segment and its neighboring associated mRNA is eventually transported to the cytoplasm, the original experiments which examined this issue (Jelinek et al., 1973) provided evidence that at least 30-40% and perhaps all of the poly(A) did appear in the cytoplasm. Recent additional kinetic analyses of labeled poly(A) accumulation in the nucleus and cytoplasm of HeLa cells again support the possibility that every poly(A) segment, and therefore the associated mRNA regions, does emerge into the cytoplasm (Puckett, Chambers and Darnell, unpublished). These new results contradict the conclusion reached by Perry et al. (1974) for L-cells. It is not known whether every poly(A) segment which enters the cytoplasm also enters polyribosomes. It appears that poly(A) terminated HnRNA is the precursor to cytoplasmic mRNA and the order of distinguishable polynucleotide regions within HnRNA is therefore the order within regions of chromosomal DNA that contain structural genes.

In contrast to the 3' poly(A) segment a number of regions of HnRNA are known not to exit to the cytoplasm, e.g., T1 RNase

resistant uridylylate rich regions, termed oligo(U), (Molloy et al., 1972) and portions of the HnRNA transcribed from repeated sites in the DNA (rapidly hybridizing RNA) (Darnell and Balint, 1970). Many of these repeated sites participate in forming intramolecular double-stranded regions in HnRNA (Jelinek and Darnell 1972; Kronenberg and Humphries, 1972; Jelinek et al., 1974). All these characteristic regions are missing in mRNA. They have been localized relative to the 3' terminus in HnRNA by first selecting poly(A) terminated molecules followed by brief alkali degradation of the selected HnRNA sample (Molloy et al., 1974). Reselection of various sized alkali-derived fragments and assays for oligo(U) and rapidly hybridizing regions reveal 90% of the oligo(U) to be over 12,000 nucleotides from the 3' terminus. The rapidly hybridizing regions (and presumably therefore double-stranded intramolecular regions) are scattered from 3,000 to 30,000 nucleotides from the 3' terminus; interspersed with the rapidly hybridizing regions are slowly hybridizing regions. Studies on sequence arrangements in DNA from *Xenopus* and sea urchins indicate that about 1/2 of the total genome is composed of alternating repeated and non-repeated regions (Davidson et al., 1973; this volume). It appears therefore that transcription of such regions gives rise to HnRNA and thus ultimately to mRNA.

Two areas of study would seem to be called for in the immediate future:

(1) Experiments to secure more conclusive evidence of mRNA derivation from HnRNA as well as to begin an understanding of what agents (presumably proteins including perhaps specific nucleases) are involved in the conversion.

(2) An inquiry into whether HnRNA containing poly(A) represents transcripts unmodified except by poly(A) addition or whether transcripts may be modified before poly(A) addition. In particular identification of a true 5' terminus on HnRNA containing structural gene transcripts is a most important order of business. Whether this can be accomplished in whole cells or must await demonstration with either purified or partially purified cell free systems remains an open question.

CONCLUSION

The aim of this paper has been to present both the difficulties and possibilities of describing the molecular details of genetic regulation in higher cells. One necessary part of such an ambitious final goal is to be able to study the production of the earliest gene transcripts to make conclusions about gene regulation. To this end studies on nuclear RNA in mammalian cells have an important continuing role to play.

REFERENCES

- Adesnik, M., M. Salditt, W. Thomas and J.E. Darnell. 1972. Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly(A) sequences and that the poly(A) has a nuclear function. *J. Mol. Biol.* 71: 21.
- Adesnik, M. and J.E. Darnell. 1972. Biogenesis and characterization of histone messenger RNA in HeLa cells. *J. Mol. Biol.* 67: 397.
- Benjamin, T. 1966. Virus-specific RNA in cells productively infected or transformed by polyoma virus. *J. Mol. Biol.* 16: 359.
- Bernhardt, D. and J.E. Darnell. 1969. tRNA synthesis in HeLa cells: A precursor to tRNA and the effects of methionine starvation on tRNA synthesis. *J. Mol. Biol.* 42: 43.
- Birnboim, H.C., R.E.J. Mitchel and N.A. Strauss. 1973. Analysis of long pyrimidine polynucleotides in HeLa cell nuclear DNA: Absence of poly(dT). *Proc. Nat. Acad. Sci. USA* 70: 2189.
- Britten, R.J. and E. Davidson. 1969. Gene regulation for higher cells: a theory. *Science* 165: 349.
- Britten, R.J. and D.E. Kohne. 1968. Repeated sequences in DNA. *Science* 161: 529.
- Burdon, R.H. and A.E. Clason. 1969. Intra-cellular location and molecular characteristics of tumour cell transfer RNA precursor. *J. Mol. Biol.* 39: 113.
- Chovnick, A. 1966. Genetic organization in higher organisms. *Proc. Roy. Soc. London B* 164: 198.
- Darnell, J.E. 1968. RNA from animal cells. *Bact. Rev.* 32: 262.
- . 1974. The origin of mRNA and the structure of the mammalian chromosome. *Harvey Lectures*, 1973-74. In press.
- Darnell, J.E. and R. Balint. 1970. The distribution of rapidly hybridizing RNA sequences in heterogeneous nuclear RNA and mRNA from HeLa cells. *J. Cell Physiol.* 76: 349.
- Darnell, J.E., L. Philipson, R. Wall and M. Adesnik. 1971a. Polyadenylic acid sequences: Role in conversion of nuclear RNA into messenger RNA. *Science* 174: 507.
- Darnell, J.E., R. Wall and R.J. Tushinski. 1971b. An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc. Nat. Acad. Sci. USA* 68: 1321.
- Darnell, J.E., W.R. Jelinek and G.R. Molloy. 1973. Biogenesis of mRNA: Genetic regulation in mammalian cells. *Science* 181: 1215.
- Davidson, E.H., B.R. Hough, C.S. Amenson and R.J. Britten. 1973. General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* 77: 1.

- Edmonds, M. and R. Abrams. 1960. Polynucleotide biosynthesis: Formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J. Biol. Chem.* 235: 1142.
- _____, _____. 1962. Isolation of a naturally occurring polyadenylate from salt thymus nuclei. *J. Biol. Chem.* 238: 1186.
- Edmonds, M.P., M.H. Vaughan and H. Nakazato. 1971. Polyadenylic acid sequences in heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. *Proc. Nat. Acad. Sci. USA* 68: 1336.
- Ellem, K.A.O. and J.W. Sheridan. 1964. Tenacious binding of the bulk of the DNA-like RNA of metazoan cells to methylated albumin columns. *Biochem. Biophys. Res. Comm.* 16: 505.
- Firtel, R.A., A. Jacobson and H.F. Lodish. 1972. Isolation and hybridization characteristic of mRNA from the cellular slime mold *D. discoideum*. *Nature New Biol.* 239: 235.
- Georgiev, G.P. and V.L. Mantieva. 1962. The isolation of DNA-like RNA and ribosomal RNA from the nucleo-chromosomal apparatus of mammalian cells. *Biochim. Biophys. Acta* 61: 153.
- Glassman, E. 1965. Genetic regulation of xanthine dehydrogenase in *D. melanogaster*. *Fed. Proc.* 24: 1243.
- Gilbert, W. 1967. The lac operator is DNA. *Proc. Nat. Acad. Sci. USA* 58: 2415.
- Greenblatt, J. and R. Schleif. 1971. Arabinose C protein: Regulation of the arabinose operon *in vitro*. *Nature New Biol.* 273: 166.
- Greenberg, T.R. and R.P. Perry. 1972. Relative occurrence of poly(A) sequences in mRNA and HnRNA of L cells as determined by poly(U) hydroxylapatite chromatography. *J. Mol. Biol.* 72: 91.
- Houssais, J.F. and G. Attardi. 1966. High molecular weight non-ribosomal type nuclear RNA and cytoplasmic messenger RNA in HeLa cells. *Proc. Nat. Acad. Sci. USA* 56: 616.
- Imaizumi, T., H. Diggelman and K. Scherrer. 1973. Demonstration of globin messenger sequences in giant nuclear precursors of mRNA of avian erythroblasts. *Proc. Nat. Acad. Sci. USA* 70: 1122.
- Jacob, F. and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3: 318.
- Jelinek, W. and J.E. Darnell. 1972. Double-stranded regions in heterogeneous nuclear RNA from HeLa cells. *Proc. Nat. Acad. Sci. USA* 69: 2537.
- Jelinek, W., M. Adesnik, M. Salditt, D. Sheiness, R. Wall, G. Molloy, L. Philipson and J.E. Darnell. 1973. Further evidence on the nuclear origin and transfer to the cytoplasm of polyadenylic acid sequences in mammalian cell RNA. *J. Mol. Biol.* 75: 515.

- Jelinek, W., G. Molloy, R. Fernandez-Munoz, M. Salditt and J.E. Darnell. 1974. Secondary structure in HnRNA: Involvement of regions from repeated DNA sites. *J. Mol. Biol.* 82: 361.
- Kates, J. 1970. Transcription of the vaccinia virus genome and the occurrence of polyadenylic acid sequences in mRNA. *Cold Spring Harbor Symp. Quant. Biol.* 35: 743.
- Kronenberg, L.H. and T. Humphries. 1972. Double-stranded RNA in sea urchin embryos. *Biochemistry* 11: 2020.
- Lee, Y., J. Mendecki and G. Brawerman. 1971. A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA components of mouse sarcoma. *Proc. Nat. Acad. Sci. USA* 68: 1331.
- Lim, L. and E.S. Canellakis. 1970. Adenine-rich polymer associated with rabbit reticulocyte messenger RNA. *Nature* 227: 710.
- Linberg, U. and J.E. Darnell. 1970. SV40-specific RNA in the nucleus and polyribosomes of transformed cells. *Proc. Nat. Acad. Sci. USA* 65: 1089.
- Maden, B.E.H. 1971. The structure and formation of ribosomes in animal cells. *Prog. Biophys. Mol. Biol.* 22: 129.
- Marzluff, G. and R.L. Metzenberg. 1968. Positive control by the *Cys-5* locus in regulation of sulfur metabolism in *Neurospora*. *J. Mol. Biol.* 33: 423.
- McAuslan, B.R. 1963. The induction and repression of thymidine kinase in the pox virus-infected HeLa cell. *Virology* 21: 383.
- McCarthy, B.J. and J.D. Duerksen. 1970. Fractionation of mammalian chromatin. *Cold Spring Harbor Symp. Quant. Biol.* 35: 621.
- McCormick, W. and S. Penman. 1968. Regulation of protein synthesis in HeLa cells: Translation at elevated temperatures. *J. Mol. Biol.* 39: 315.
- Melli, M. and R. Pemberton. 1972. New method of studying the precursor product relationship between high molecular weight RNA and messenger RNA. *Nature New Biol.* 236: 172.
- Mendecki, J.W., E. Lee and G. Brawerman. 1972. Characteristics of the polyadenylic acid segment associated with messenger ribonucleic acid in mouse sarcoma 180 ascites cells. *Biochemistry* 11: 5.
- Miller, O.L., B.A. Hamkalo and C.A. Thomas. 1970. Visualization of bacterial genes in action. *Science* 169: 392.
- Molloy, G.R. and J.E. Darnell. 1973. Characterization of the polyadenylic acid regions and the adjacent nucleotides in heterogeneous nuclear ribonucleic acid and messenger ribonucleic acid from HeLa cells. *Biochemistry* 12: 2324.
- Molloy, G.R., W.L. Thomas and J.E. Darnell. 1972. Occurrence of uridylylate-rich oligonucleotide regions in heterogeneous nuclear RNA of HeLa cells. *Proc. Nat. Acad. Sci. USA* 69: 3684.

- Molloy, G.R., W. Jelinek, M. Salditt and J.E. Darnell. 1974. Arrangement of specific oligonucleotides with poly(A) terminated HnRNA molecules. *Cell* 1: 43.
- Pagoulatos, G.N. and J.E. Darnell. 1970. A comparison of the heterogeneous nuclear RNA of HeLa cells in different periods of the cell growth cycle. *J. Cell. Biol.* 44: 476.
- Penman, S., K. Scherrer, Y. Becker and J.E. Darnell. 1963. Polyribosomes in normal and poliovirus-infected HeLa cells and their relationship to messenger RNA. *Proc. Nat. Acad. Sci. USA* 49: 654.
- Penman, S., I. Smith and E. Holtzman. 1966. Ribosomal RNA synthesis and processing in a particulate site in the HeLa cell nucleus. *Science* 154: 786.
- Penman, S., M. Rosbash and M. Penman. 1970. Messenger and heterogeneous nuclear RNA in HeLa cells. Differential inhibition by cordycepin. *Proc. Nat. Acad. Sci. USA* 67: 1878.
- Perry, R.P. 1962. The cellular sites of synthesis of ribosomal and 4S RNA. *Proc. Nat. Acad. Sci. USA* 48: 2179.
- Perry, R.P. and D.E. Kelley. 1973. Messenger RNA turnover in mouse L cells. *J. Mol. Biol.* 79: 681.
- Perry, R.P., J.R. Greenberg and K.D. Tartof. 1970. Transcription of ribosomal, heterogeneous nuclear, and messenger RNA in eukaryotes. *Cold Spring Harbor Symp. Quant. Biol.* 35: 577.
- Perry, R.P., D.E. Kelley and J. LaTorre. 1974. Synthesis and turnover of nuclear and cytoplasmic polyadenylic acid in mouse L cells. *J. Mol. Biol.* 82: 315.
- Philipson, L., R. Wall, R. Glickman and J.E. Darnell. 1971. Addition of polyadenylate sequences to virus-specific RNA during adenovirus replication. *Proc. Nat. Acad. Sci. USA* 68: 2806.
- Sambrook, J., H. Westphal, P.R. Srinivasan and R. Dulbecco. 1968. The integrated state of viral DNA in SV40-transformed cells. *Proc. Nat. Acad. Sci. USA* 60: 1288.
- Scherrer, K., H. Latham and J.E. Darnell. 1963. Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. *Proc. Nat. Acad. Sci. USA* 49: 240.
- Scherrer, K., G. Spohr, N. Granbault, C. Morel, J. Grosclaude and C. Chezzi. 1970. Nuclear and cytoplasmic messenger-like RNA and their relation to the active messenger RNA in polyribosomes of HeLa cells. *Cold Spring Harbor Symp. Quant. Biol.* 35: 539.
- Scherrer, K. and L. Marcaud. 1965. Remarques sur les ARN messagers polycistronique dans les cellules animales. *Bull. Soc. Chim. Biol.* 47: 1697.
- Shearer, R.N. and R.J. McCarthy. 1967. Evidence for ribonucleic acid molecules restricted to the cell nucleus. *Biochemistry* 6: 283.

- Sibatani, E., S.R. deKloet, V.G. Allfrey and A.E. Mirsky. 1962. Isolation of a nuclear RNA fraction resembling DNA in its base composition. *Proc. Nat. Acad. Sci. USA* 48: 471.
- Singer, R.H. and S. Penman. 1973. Messenger RNA in HeLa cells: Kinetics of formation and decay. *J. Mol. Biol.* 78: 321.
- Soeiro, R., H.C. Birnboim and J.E. Darnell. 1966. Rapidly labeled HeLa cell nuclear RNA. II. Base composition and cellular localization of a heterogeneous RNA fraction. *J. Mol. Biol.* 19: 362.
- Soeiro, R. and J.E. Darnell. 1970. A comparison between HnRNA and polysomal mRNA in HeLa cells by RNA-DNA hybridization. *J. Cell Biol.* 44: 467.
- Tompkins, G.M., T. Gelehrger, D. Granner, D.W. Martin, H. Samuels and E.B. Thompson. 1969. Control of specific gene expression in higher organisms. *Science* 166: 1474.
- Tonegawa, S., G. Walter, A. Bernardini and R. Dulbecco. 1970. Transcription of the SV40 genome in transformed cells and during lytic infection. *Cold Spring Harbor Symp. Quant. Biol.* 35: 823.
- Wall, R., J. Weber, Z. Gage and J.E. Darnell. 1973. The production of viral messenger RNA in adenovirus transformed cells by the post-transcriptional processing of heterogeneous nuclear RNA containing viral and cell sequences. *J. Virol.* 11: 953.
- Williamson, R., C. Drewienkiewicz and J. Paul. 1973. Globin mRNA sequences in high molecular weight RNA from embryonic mouse liver. *Nature New Biol.* 241: 66.
- Yosikawa-Fukada, M., T. Fukada and Y. Kawade. 1965. Characterization of rapidly labeled ribonucleic acid of animal cells in culture. *Biochim. Biophys. Acta* 103: 383.
- Zubay, G., D. Schwartz and J. Beckwith. 1970. Mechanism of activation of catabolite-sensitive genes: A positive control system. *Proc. Nat. Acad. Sci. USA* 66: 104.

Oligothymidylic acid tracts in mouse DNA

Jonathan N. Mansbridge, George D.
Birnie, Steve Humphries, Som
Atisontornkul and John Paul

When pyrimidine tracts derived from mouse DNA are passed through a column of poly-[A]-Sepharose, fragments amounting to 0.051% of the total DNA are bound to the column and can be eluted with formamide. On electrophoresis in formamide gels, these fragments are found to be heterogeneous, ranging in size from 10 to greater than 100 residues. Pyrimidine tracts prepared in this way contain a substantial proportion of cytosine residues in addition to thymidine sequences. Fragments containing predominantly thymidine residues have been isolated by annealing depurinated DNA to poly-[A] and digesting with nuclease S1. About 65% of the material is rendered acid soluble. The remaining fragments, which may contain a few cytosine residues, move as a comparatively homogeneous band on electrophoresis. The fragments which bind to poly-[A]-Sepharose may also be eluted by means of a temperature gradient. The elution takes place over a range of temperatures up to 53°, with a maximum at 31°, which corresponds to the melting temperature of an oligo-[T] of about 11 residues. It is unlikely that these sequences function as the template for the synthesis of the poly-[A] segment of mRNA nor as coding sequences. The occurrence of the sequences is discussed in terms of control functions and protection against damage by ultraviolet light.

Estimates of the binding of poly-[A] and poly-[U] to DNA have suggested that substantial numbers of sequences containing oligo-[dT] and oligo-[dA] occur in mammalian genomes (Shenkin and Burdon, 1972; Williamson and Morrison, 1971). Studies on these sequences were originally carried out to determine whether they might form a template for the synthesis of the poly-[A] segment of mRNA. The available evidence, both from kinetic and from sequence studies, indicates that these sequences arise from post-transcriptional addition (Darnell et al., 1971; Birnboim et al., 1973), but the observation that the powerful chain terminating mutagen dideoxy-thymidine appears to be preferentially incorporated into long pyrimidine tracts has increased the interest in these sequences (Kidson et al., 1975). These sequences have been investigated further by isolation of pyrimidine tracts which contain sufficient runs of thymidine to permit them to bind to a poly-[A]-Sepharose affinity column.

METHODS

Isolation of oligo-[dT] tracts DNA was depurinated by the method of Burton (1967) and freeze dried. Pyrimidine tracts containing substantial oligo-[dT] regions were extracted by binding them to a poly-[A]-Sephadex column. Poly-[A] was bound to Sephadex by means of cyanogen bromide (Wagner et al., 1971). Columns with a total volume of 1 ml were equilibrated with NETS buffer (0.1 M NaCl, 0.001 M EDTA, 0.01 M Trizma, pH 7.4, 0.2% SDS) (Adesnik et al., 1972), and apurinic acid applied in the same buffer. Unbound nucleotides were removed by washing with at least 20 ml NETS. Tracts were eluted either with 3 ml 90% formamide, 0.01 M Trizma, pH 7.4, 0.001 M EDTA, 0.2% SDS or by increasing the temperature.

Electrophoresis Polyacrylamide gel electrophoresis was carried out in 17% gels in barbital buffer in 98% formamide at an apparent pH of 9.0.

RESULTS

The binding of pyrimidine tracts from DNA labelled with tritiated thymidine has been measured for mouse and *Escherichia coli* DNA. In the case of mouse, 0.14% of such label can be bound to a poly-[A]-Sephadex column. This contrasts with a value of 0.0045% for *E. coli*. If the mouse tracts are annealed with poly-[A] and the hybrids treated with S1 nuclease, 35% of the poly-[A]-Sephadex binding fragments become acid soluble and the fraction of thymidine binding to the column is 0.051%. This corresponds to 0.015% of mouse DNA. The existence of such tracts in calf thymus and human DNA has also been demonstrated.

Temperature elution profile of poly-[T] tracts The elution of mouse pyrimidine tracts from a column of poly-[A]-Sephadex is set out in Figure 1. In this experiment, the column was washed at 19° until the activity in the samples was 0.3% of the activity remaining on the column. The temperature was then raised by steps of 2°. The temperature at which 50% of the material had eluted was 31°. Columns of this type have been calibrated using oligo-[dT] sequences of known chain length, and the melting temperature of the pyrimidine tracts corresponds to an average chain length of 11 residues (Figure 1).

Electrophoresis of oligo-[dT] containing pyrimidine tracts On electrophoresis, the pyrimidine tracts which bind to poly-[A]-Sephadex are found to possess a variety of molecular weights ranging from about 11 residues up to several hundreds. After annealing with poly-[A] and treatment with S1 nuclease, the remaining oligo-[dT] tracts are found, on electrophoresis, in a rather sharp peak of lower molecular weight. These results confirm that many of the oligo-[dT] containing tracts also contain a considerable proportion of cytosine elsewhere in the structure.

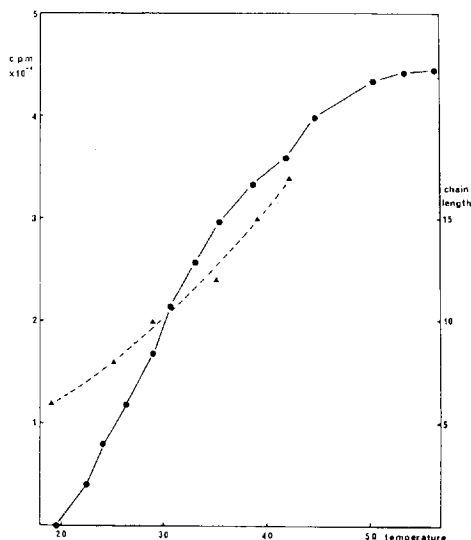


Figure 1. Thermal elution profile of mouse pyrimidine tracts from a poly-[A]-Sephadex column. (●—●) — elution profile of pyrimidine tracts. (▲—▲) — elution temperatures of oligo-[dT] of known chain length.

DISCUSSION

The results obtained in this study show that oligo-[dT] tracts of substantial length occur in the DNA of several mammals. The temperature elution profile of these tracts indicates that they behave as though they are about 11 residues long. It cannot be excluded, however, that they may contain an occasional cytosine residue, but for an oligo-[dT] tract containing one cytosine residue to behave in the same way it would have to contain about 17 residues overall. The average length of the oligo-[dT] tracts would then be eight residues.

From the proportion of oligo-[dT] tracts in mouse DNA and their average size it can be calculated that the haploid genome contains about 5×10^4 such tracts. This is about 100 times the frequency that would be expected on a random basis.

The possible function of these sequences is obscure. It is clearly unlikely that they form coding regions in structural genes as they would result in runs of lysine or phenylalanine at a greater frequency than has been observed. They are present in sufficient numbers to act as templates for the synthesis of the poly-[A] segment of mRNA but they are much too short, such sequences ranging from 50 residues in globin mRNA (Mansbridge et al., 1974) to several hundreds in other cases (Molloy and Darnell, 1973). It is clear

that the major part of these sequences is added post-transcriptionally, as indicated by kinetic evidence (Darnell et al., 1971). However, the possibility that oligo-[dT] sequences may lie at the distal ends of genes and act as primers for the addition of further adenylic acid residues is not excluded. Indeed, such sequences might act as terminators for RNA synthesis.

A further possibility is that they are concerned with protecting DNA from damage by ultraviolet irradiation. It has been reported by Shavranovskaya et al. (1973) that thymine dimers formed as a result of irradiation occur in clusters, and Brunk (1973) has found that in the DNA of *Tetrahymena pyriformis* they are formed preferentially in pyrimidine tracts. It is possible, therefore, that the thymidylic acid sequences may act as traps for UV excitation of the DNA molecule and concentrate dimer formation away from the structural gene and control regions of the molecule.

The most attractive possibility is that they form the template for the oligo-[U] region of hnRNA (Molloy et al., 1972). In this regard they form an identifiable sequence marker in DNA.

REFERENCES

- Adesnik, M., M. Salditt, W. Thomas and J.E. Darnell. 1972. Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly[A] sequences and that poly[A] has a nuclear function. *J. Mol. Biol.* 71: 21.
- Birnboim, H.C., R.E.J. Mitchell and N.A. Straus. 1973. Analysis of long pyrimidine tracts in HeLa nuclear DNA. Absence of polydeoxythymidylate. *Proc. Nat. Acad. Sci. USA* 70: 2189.
- Brunk, C.F. 1973. Distribution of dimers in ultraviolet irradiated DNA. *Nature New Biol.* 241: 74.
- Burton, M. 1967. Preparation of apurinic acid and oligodeoxyribonucleotides with formic acid and diphenylamine. *In Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 12A, p. 222. Academic Press, New York.
- Darnell, J.E., L. Philipson, R. Wall and M. Adesnik. 1971. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science* 174: 507.
- Kidson, C., P.C. Chen, T. Edwards and J. Mansbridge. 1975. Chromosome functional organization determined by replication mapping. This volume.
- Mansbridge, J.N., J.A. Crossley, W.G. Lanyon and R. Williamson. 1974. The polyadenylic acid region of mouse globin mRNA. *Europ. J. Biochem.* In press.
- Molloy, G.R., W.L. Thomas and J.E. Darnell. 1972. Occurrence of uridylylate-rich oligonucleotide regions in heterogeneous nuclear RNA of HeLa cells. *Proc. Nat. Acad. Sci. USA* 69: 3684.

- Molloy, G.R. and J.E. Darnell. 1973. Characterization of the poly(adenylic acid) regions and the adjacent nucleotides in heterogeneous nuclear ribonucleic acid and messenger ribonucleic acid from HeLa cells. *Biochemistry* 12: 2324.
- Shafranovskaya, N.N., E.N. Trifonov, Yu. S. Lazurkin and M.D. Frank-Kamenetskii. 1973. Clustering of thymine dimers in ultraviolet irradiated DNA and the long-range transfer of electronic excitation along the molecule. *Nature New Biol.* 241: 58.
- Shenkin, A. and R.H. Burdon. 1972. Deoxyadenylate-rich sequences in mammalian DNA. *Fed. Europ. Biochem. Soc. Letters* 22: 157.
- Williamson, R. and M.R. Morrison. 1971. The isolation and DNA/RNA hybridization of messenger RNA for globin. *Ser. Haematol.* IV: 23.
- Wagner, A.F., R.L. Bugianesi and T.Y. Shen. 1971. Preparation of Sepharose-bound poly(rI:rC). *Biochem. Biophys. Res. Comm.* 45: 184.

Properties of the mRNA for feather keratin: mRNA transcribed from a family of repeated genes

David J. Kemp, Ian D. Walker,
Geoffrey A. Partington and George
E. Rogers

Keratin mRNA was isolated in apparently pure form from 14-day embryonic chick feather polysomes and contains long untranslated sequence(s). Hybridization studies with complementary DNA prepared from this mRNA indicated that there are from 100-300 keratin genes in the chick genome compared with the figure of at least 30 genes determined from protein chemical studies.

Studies on keratin synthesis in the developing embryonic chick feather have led to the conclusion that feather keratin chains are the products of multiple homologous non-allelic genes. The keratin genes can be regarded as a family of moderately repetitious non-identical sequences. In this paper, our studies on feather keratin proteins, on the purification and properties of feather keratin mRNA and on the number of keratin genes are reviewed.

RESULTS AND DISCUSSION

Keratin synthesis during development of the embryonic chick feather
Polyacrylamide gel electrophoresis of total cellular proteins from 11-19-day embryonic chick feathers (Kemp et al., 1974) revealed a dramatic change in the protein pattern over this period (Figure 1). In the 11- and 12-day feather extracts the major feather keratin chains (bands $\beta 2 - \beta 5$; Kemp and Rogers, 1972) were present only as traces. These bands became the most abundant protein species in the feathers by day 15.

Initial characterization by polyacrylamide gel electrophoresis of the keratin chains present in the 21-day (i.e. newly hatched) embryonic chick feather demonstrated the presence of at least 10 components and additional keratin chains were present in adult tissues (Kemp, 1972; Kemp and Rogers, 1972). DEAE cellulose chromatography of 21-day feather proteins followed by gel electrophoresis indicated that there were at least 19-25 distinct keratin chains present (Walker, 1974).

Sequence studies on embryonic feather keratins (Walker, 1974) and adult scale keratins demonstrated the presence of multiple amino

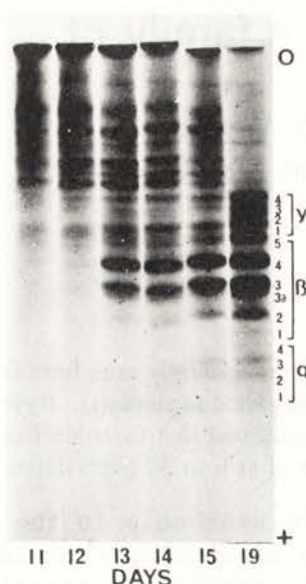


Figure 1. Polyacrylamide gel electrophoresis at pH 9.5 of proteins from feathers during embryonic development. Total proteins were extracted from feathers from 11-19 day chick embryos by reduction and carboxymethylation (Kemp et al., 1974). Samples of the proteins were subjected to polyacrylamide gel electrophoresis at pH 9.5. O: origin. +: anode. Bands $\alpha 1 - \gamma 4$ (Kemp and Rogers, 1972) are indicated.

acid substitutions in the primary structure (Figure 2), strongly supporting the hypothesis that the large number of keratin chains are the products of multiple homologous genes. The majority of the multiple keratin genes are nonallelic (Kemp, 1972).

We conclude from the protein chemical studies of both embryonic and adult tissues that there are at least 30 nonallelic homologous keratin genes in the chick genome.

Identification and purification of feather keratin mRNA A 12S RNA isolated from 14-day embryonic feather polysomes programmed the synthesis of keratin chains in a cell-free system from rabbit reticulocytes (Partington et al., 1973).

Fractionation of 14-day feather polysomal RNA on unmodified cellulose (Kitos et al., 1972; Schutz et al., 1972; DeLarco and Guroff, 1973) demonstrated a bound 12S RNA fraction (Figure 3a) which was active in keratin synthesis (Partington et al., 1973). When RNA was obtained from 12-day feather polyribosomes, before the major onset of keratin synthesis, the 12S RNA peak was absent from the cellulose-bound fraction (Figure 3b).

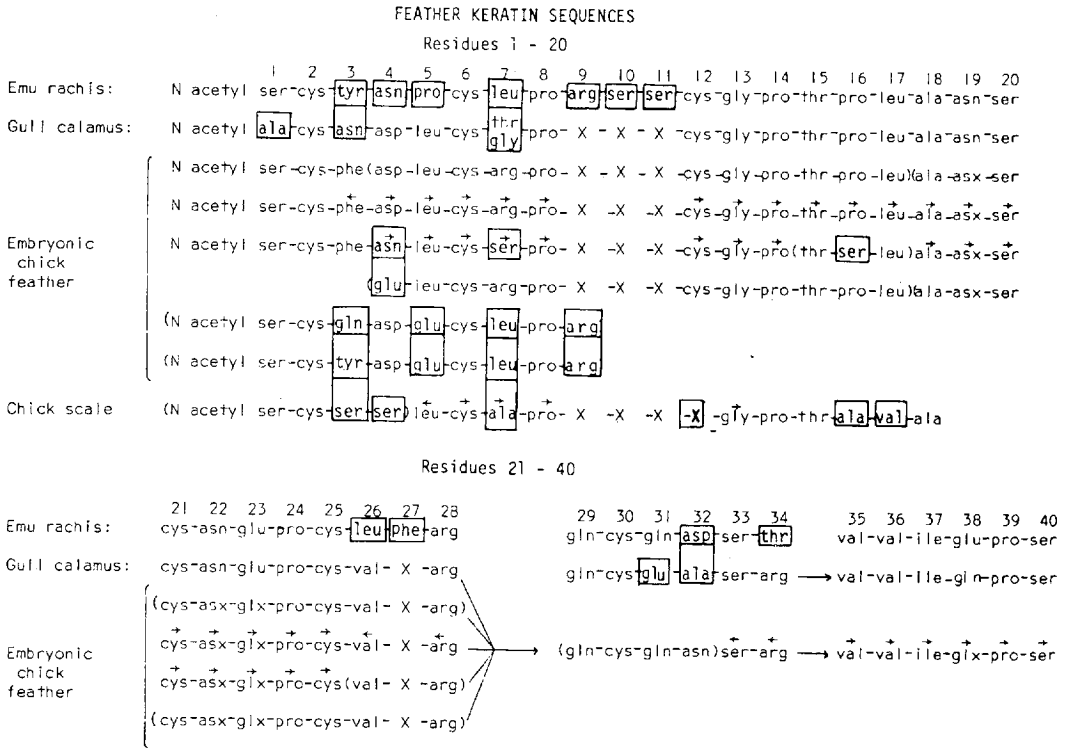


Figure 2. N-terminal amino acid sequences of feather keratins. The sequences of the emu and gull keratins are from O'Donnell (1973) and O'Donnell and Inglis (1974). The chick feather keratin sequences were determined on tryptic peptides of 21-day embryonic feather keratins (Walker, 1974) and have been aligned to maximise homology with the emu and gull sequences. Amino acid sequences in parenthesis have been inferred from amino acid compositions and homology considerations. →; amino acid residue identified by dansylation ←; amino acids identified after carboxypeptidase digestion of the peptide. X; amino acid deletion. Amino acids surrounded by a box differ from the top.

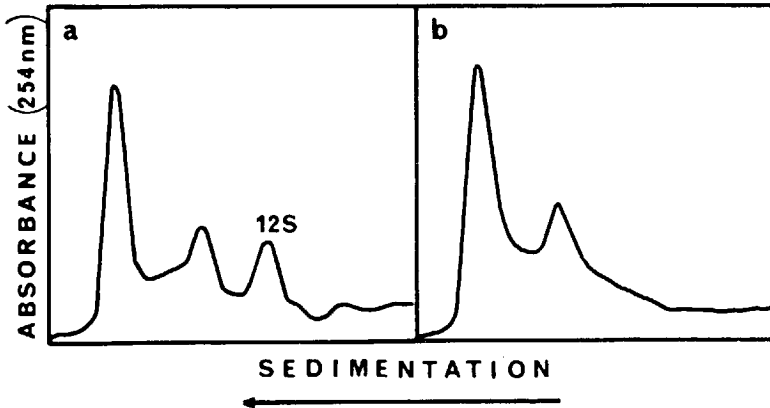


Figure 3. Sucrose gradient centrifugation of the cellulose-bound fraction of 12- and 14-day feather polysomal RNA. Polysomal RNA was prepared by phenol extraction from 12- and 14-day feather polysomes as described (Partington et al., 1973) and fractionated by cellulose chromatography (a) 14-day cellulose-bound RNA. (b) 12-day cellulose-bound RNA.

Chantrenne et al. (1967) demonstrated that rabbit globin mRNA could be released as a protein complex (mRNP) by treatment of polyribosomes with EDTA. The feather mRNP released by EDTA treatment contained a major RNA species sedimenting at 12S (Figure 4a), and small amounts of RNA sedimenting at 5, 7, 9, 14 and 18S. The cellulose-bound fraction of the RNA from feather RNP was selectively enriched in the 12S and 9S RNA species (Figures 4b and c).

Purity of feather keratin mRNA 12S feather keratin mRNA obtained either by cellulose chromatography or from mRNP contained only 1 band of apparent MW 250,000 (Figure 5), when electrophoresed on polyacrylamide gels in the presence of 98% formamide (Staynov et al., 1972; Gould and Hamlyn, 1973). Its translation in the wheat embryo

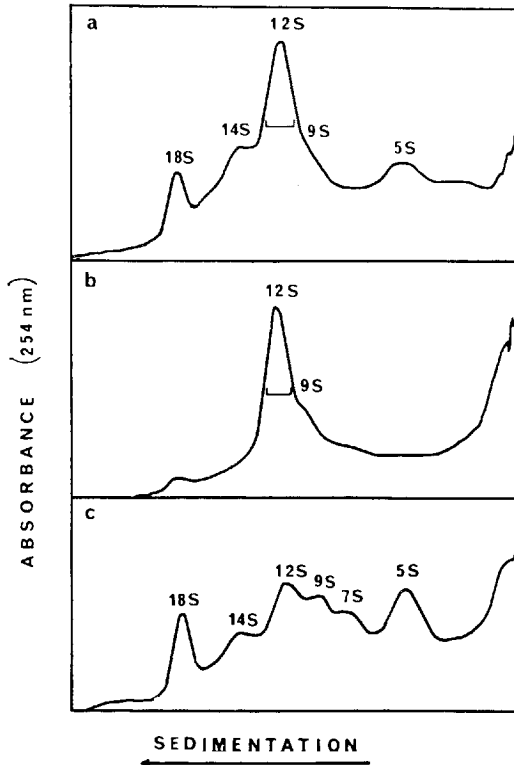


Figure 4. Sucrose gradient centrifugation of RNA fractions obtained from mRNPs, after dissociation with SDS. (a) feather mRNP; (b) cellulose-bound fraction of RNA from mRNP; RNA was isolated by phenol extraction of mRNP and fractionated on cellulose. (c) As for (b), but the non-bound fraction.

cell-free system (Efron and Marcus, 1973; Roberts and Paterson, 1973; Shib and Kaesberg, 1973), yielded about 95% of the incorporated radioactivity migrating as one band and co-electrophoresing with feather keratin on SDS-polyacrylamide gels.

Fractionation of the labelled material on polyacrylamide gels at pH 7.5 demonstrated that the labelled material co-electrophoresed with keratin bands β_2 - β_5 .

In summary then, keratin chains are the major product of 14-day feathers but not of 12-day feathers and a 12S RNA is present in 14-day but not in 12-day feather polyribosomes. This RNA is released as a mRNP complex from polyribosomes by EDTA treatment, binds selectively to cellulose, migrates as a discrete band during polyacrylamide gel electrophoresis, codes only for feather keratin chains in the wheat embryo cell-free system and is as active in translation as rabbit globin mRNA. We conclude that this RNA is greater than 95% pure keratin mRNA.



Figure 5. Polyacrylamide gel electrophoresis of RNA in the presence of 98% formamide. (a) Cellulose-bound 12S feather keratin mRNA, prepared as shown in Figure 3a. (b) As for (a), but the non-bound 12S fraction. (c) 12S feather keratin mRNA from SDS dissociation of RNP, prepared as shown in Figure 4a. (d) Globin mRNA, prepared by SDS dissociation of RNP. (e) As for (c), but with the addition of *E. coli* 16S and 23S rRNA, and rabbit reticulocyte 18S rRNA and 4-5S RNA. (f) As for (e), but with the addition of globin mRNA as in (d). The globin mRNA preparation shown in (f) contained only one band, while that in (d) contained two bands.

This keratin mRNA has a MW of about 250,000, and thus is 800 bases long. Assuming it is monocistronic (Partington, 1974) about 300 bases are required to code for keratin chains, and so about 500 bases in the non-coding regions is indicated. Untranslated sequences in eukaryotic mRNA appear to be of general occurrence (Lim and Canellakis, 1970; Berns et al., 1972; Gould and Hamlyn, 1973; Schechter, 1973). Since the keratin mRNA apparently codes for all of the embryonic feather chains, it is significant that only one discrete band of mRNA was evident on formamide gels. The length of the untranslated sequence(s) must have been conserved during the evolutionary divergence of the keratin genes.

Synthesis of DNA complementary to feather keratin mRNA Keratin mRNA acted as an efficient template for the synthesis of a DNA complement (cDNA) in the presence of the purified polymerase from avian myeloblastosis virus (Kemp and Rogers, unpublished results) as reported

for several poly-A containing mRNAs (Kacian et al., 1972; Ross et al., 1972; Verma et al., 1972). Keratin mRNA was about 25-50% as active a template as rabbit globin mRNA and transcription was oligo dT-dependent. Fractionation of keratin cDNA and globin cDNA on neutral sucrose gradients indicated that the keratin cDNA was of greater maximum size (8-10S) than the globin cDNA (7-8S) but was more heterogeneous.

Hybridization of keratin cDNA to keratin mRNA The rate of hybridization of keratin mRNA to keratin cDNA was lower than that of globin mRNA to globin cDNA by a factor of about 30-35 (Figure 6), when assayed by resistance of the hybrids to nuclease S1 (Ando, 1966; Vogt, 1973).

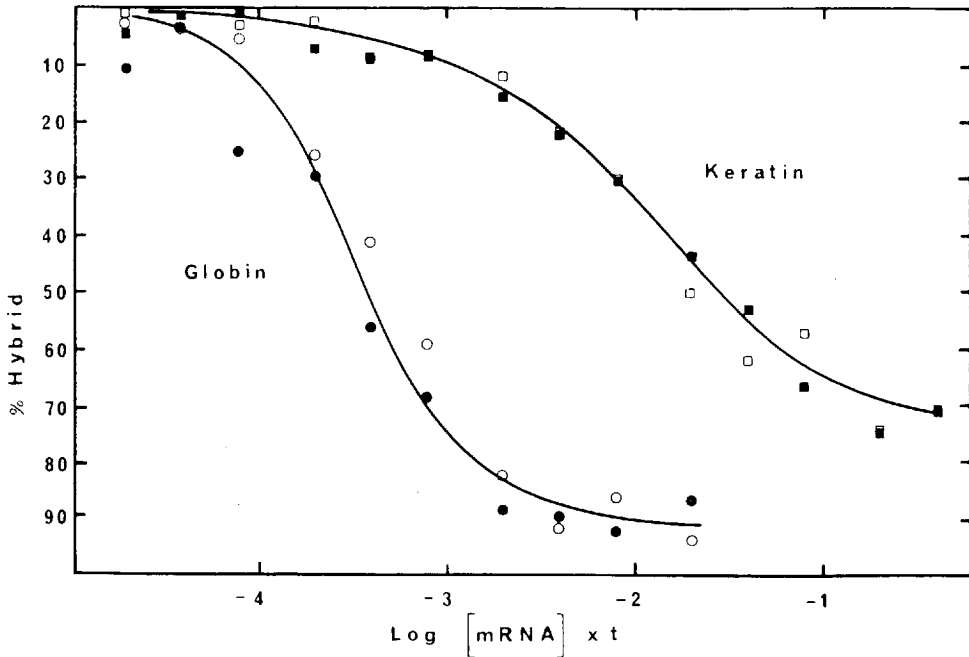


Figure 6. Hybridization of keratin mRNA and globin mRNA to cDNA. Hybridizations were at 60° for 4 hours in 0.18 M NaCl, 0.01 M Tris/HCl, 0.001 M EDTA, 0.05% SDS, pH 7.0. The reaction mixtures (25 ul) contained, 600 cpm of keratin cDNA or globin cDNA, and varying concentrations of keratin mRNA or globin mRNA. After incubation, the reaction mixtures were incubated at 45° for 30 min with and without nuclease S1 (4 units)(Vogt, 1973), TCA precipitated and counted. Results are expressed as the percent cpm resistant to nuclease S1 digestion. ■ — ■: keratin mRNA-cDNA. □ — □: keratin mRNA-cDNA, independent experiment with a different batch of keratin mRNA. ● — ●, ○ — ○: globin mRNA-cDNA, two independent experiments.

The rate of hybridization of globin mRNA to globin cDNA observed was similar to published values (Bishop and Rosbach, 1973; Kacian et al., 1973; Gummerson and Williamson, 1974). Globin mRNA is a two-component system since cross-hybridization of the α - and β -chain mRNAs with the cDNAs does not occur (Kacian et al., 1973; Housman et al., 1973). The observed rate of hybridization of keratin mRNA to keratin cDNA is therefore about 60 times slower than that expected for a one-component system. The kinetic complexity of the keratin mRNA determined by this procedure suggests that it contains a large number of discrete keratin mRNA species. However, mismatching of cDNA transcribed from one species of keratin mRNA and hybridizing with other species of keratin mRNA would be expected to lower the observed rate of the reaction (Southern, 1970; Bonner et al., 1973). The slow hybridization of keratin mRNA to cDNA and low stability to S1 of the hybrids (Figure 6) may therefore be partly explained by mismatching (Kacian et al., 1973; Housman et al., 1973; Gummerson and Williamson, 1974) although its exact contribution to the overall rate cannot be determined from the present data.

Determination of the reiteration frequency of the genes for feather keratin Under conditions of vast DNA excess (Melli et al., 1971; Bishop and Rosbash, 1973), the observed rate of reannealing of keratin cDNA was from 5 to 10 times faster than the unique fraction (Britten and Kohne, 1968; Laird, 1971) of chick DNA (Figure 7) when assayed by the S1 procedure and from 20 to 50 times faster when assayed on hydroxyapatite.

We conclude from these results that there are about 100-300 keratin genes in the haploid chick genome. Studies on the T_m of the cDNA-DNA hybrids should allow a better estimate of the number of keratin genes (Bonner et al., 1973).

CONCLUSION

Feather keratin mRNA can be obtained in an active and essentially pure state, and is apparently a family of closely homologous mRNAs. The hybridization studies establish that there are many keratin genes in the chick genome and whilst the exact number cannot be stated, we favour the interpretation of about 100-300. These can be considered to be a set of moderately reiterated but non-identical DNA sequences and therefore the feather keratin gene can be regarded as apparently enriched 100-300-fold compared to other unique sequences in the chick genome. This feature presents an obvious advantage for the attempted isolation of 'unique' DNA sequences. If they are clustered on isolated DNA molecules, the system may have significant value for the study of the organization of genes in the avian genome and gene organization in eukaryotic systems in general.

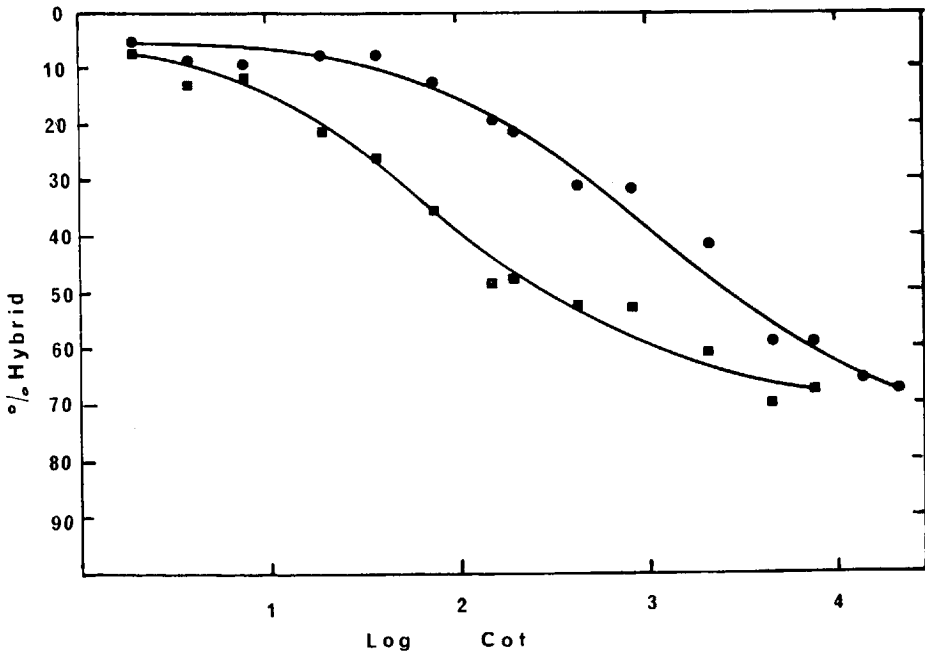


Figure 7. Renaturation of keratin cDNA and ^{125}I -DNA with chick erythrocyte nuclei DNA. Chick erythrocyte DNA was prepared and sheared to fragments of mean length $0.16\ \mu$. This fragmented DNA was iodinated (Teraba and McCarthy, 1973). Mixtures of the DNA containing 20,000 cpm/ml of either ^3H -cDNA or ^{125}I -DNA at a final DNA concentration of 6.66 mg/ml in the hybridization solution (see Figure 6) were prepared and incubated to the C_{0t} values shown. The samples were then assayed for nuclease S1 resistance. ●—● ^{125}I -DNA. ■—■, keratin cDNA.

ACKNOWLEDGMENTS

We are indebted to Dr J.R.E. Wells of this Department for donating purified AMV polymerase and criticising the manuscript. AMV was generously donated (to J.R.E.W.) by the N.I.H. cancer program. We thank Ms L. Bissell for technical assistance and Ms K. Murray for typing the manuscript. This study was supported by a grant from the Australian Research Grants Committee (to G.E.R.).

REFERENCES

- Ando, T. 1966. A nuclease specific for heat-denatured DNA isolated from a product of *Aspergillus oryzae*. *Biochim. Biophys. Acta* 114: 158.

- Berns, A.J.M., G. Strous and H. Bloemendal. 1972. Heterologous *in vitro* synthesis of lens α -crystallin polypeptide. *Nature New Biol.* 236: 7.
- Bishop, J.O. and M. Rosbash. 1973. Reiteration frequency of duck haemoglobin genes. *Nature New Biol.* 241: 204.
- Britten, R.J. and D.E. Kohne. 1968. Nucleotide sequence repetition in DNA. *Carnegie Institution of Washington, Yearbook* 67: 78.
- Bonner, T.I., D.J. Brenner, B.R. Neufield and R.J. Britten. 1973. Reduction in the rate of DNA reassociation by sequence divergence. *J. Mol. Biol.* 81: 123.
- Chantrenne, H., A. Burny and G. Marbaix. 1967. The search for the messenger RNA of haemoglobins. In *Progress in Nucleic Acid Research* (J.N. Davidson and W.E. Cohn, ed.) Vol.7, p.173. Academic Press, New York.
- Delarco, J. and G. Guroff. 1973. The binding of RNA to various celluloses. *Biochem. Biophys. Res. Comm.* 50: 486.
- Efron, D. and A. Marcus. 1973. Efficient synthesis of rabbit globin in a cell-free system from wheat embryo. *Fed. Europ. Biochem. Soc. Letters* 33: 23.
- Gould, M.J. and P.H. Hamlyn. 1973. The molecular weight of rabbit globin mRNAs. *Fed. Europ. Biochem. Soc. Letters* 30: 301.
- Gummerson, K.S. and R. Williamson. 1974. Sequence divergence of mammalian globin messenger RNA. *Nature* 247: 265.
- Housman, D., B.G. Forget, A. Skoultchi and E.J. Benz. 1973. Quantitative deficiency of chain-specific globin messenger ribonucleic acids in the thalassemia syndromes. *Proc. Nat. Acad. Sci. USA* 70: 1809.
- Kacian, D.L., S. Spiegelman, A. Bank, M. Terada, S. Metafora, L. Dow and P.A. Marks. 1972. *In vitro* synthesis of DNA components of human genes for globins. *Nature New Biol.* 235: 167.
- Kacian, D.L., R. Gambino, L.W. Dow, E. Grossbard, C. Natta, F. Ramirez, S. Spiegelman, P.A. Marks and A. Bank. 1973. Decreased globin mRNA in thalassemia detected by molecular hybridization. *Proc. Nat. Acad. Sci. USA* 70: 1886.
- Kemp, D.J. 1972. Feather keratin synthesis. Ph.D. thesis, University of Adelaide.
- Kemp, D.J. and G.E. Rogers. 1972. Differentiation of avian keratinocytes. Characterization and relationships of the keratin proteins of adult and embryonic feathers and scales. *Biochemistry* 11: 969.
- Kemp, D.J., P.Y. Dyer and G.E. Rogers. 1974. Keratin synthesis during development of the embryonic chick feather. *J. Cell. Biol.* 62: 114.

- Kitos, P.A., G. Saxon and H. Amos. 1972. The isolation of polyadenylate with unreacted cellulose. *Biochem. Biophys. Res. Comm.* 47: 1426.
- Laird, C.D. 1971. Chromatid structure: Relationships between DNA content and nucleotide sequence diversity. *Chromosoma* 32: 378.
- Lim, L. and E.S. Canellakis. 1970. Adenine-rich polymer associated with rabbit reticulocyte mRNA. *Nature* 227: 710.
- Melli, M., C. Whitfield, K.V. Rao, M. Richardson and J.O. Bishop. 1971. DNA-RNA hybridization in vast DNA excess. *Nature* 231: 8.
- O'Donnell, I.J. 1973. The complete amino acid sequence of a feather keratin from emu. *Aust. J. Biol. Sci.* 26: 415.
- O'Donnell, I.J. and A.S. Inglis. 1974. The amino acid sequence of a feather keratin from silver gull (*Larus novae-hollandiae*) - comparison with one from emu feather (*Dromaius novae-hollandiae*). *Aust. J. Biol. Sci.* In press.
- Partington, G.A. 1974. Studies on feather keratin synthesis *in vitro*. Ph.D. thesis, University of Adelaide.
- Partington, G.A., D.J. Kemp and G.E. Rogers. 1973. Isolation of feather keratin mRNA and its translation in a rabbit reticulocyte cell-free system. *Nature New Biol.* 246: 33.
- Roberts, B.E. and B.M. Paterson. 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proc. Nat. Acad. Sci. USA* 70: 2330.
- Ross, J., H. Aviv, E. Scolnick and P. Leder. 1972. *In vitro* synthesis of DNA complementary to purified rabbit globin mRNA. *Proc. Nat. Acad. Sci. USA* 69: 264.
- Schechter, I. 1973. Biologically and chemically pure mRNA coding for a mouse immunoglobulin L-chain prepared with the aid of antibodies and immobilized oligothymidine. *Proc. Nat. Acad. Sci. USA* 70: 2256.
- Schutz, G., M. Beato and P. Feigelson. 1972. Isolation of eukaryotic mRNA on cellulose and its translation *in vitro*. *Biochem. Biophys. Res. Comm.* 49: 680.
- Shih, D.S. and P. Kaesberg. 1973. Translation of Brome Mosaic Viral RNA in a cell-free system derived from wheat embryo. *Proc. Nat. Acad. Sci. USA* 70: 1799.
- Southern, E.M. 1970. Base sequence and evolution of guinea-pig α -satellite DNA. *Nature* 227: 794.
- Staynov, D.Z., J.C. Pinder and W.B. Gratzner. 1972. Molecular weight determinations of nucleic acids by gel electrophoresis in non-aqueous solution. *Nature New Biol.* 235: 108.

- Tereba, A. and B.J. McCarthy. 1973. Hybridization of ^{125}I -labelled ribonucleic acid. *Biochemistry* 12: 4675.
- Verma, I.M., G.F. Temple, H. Fan and D. Baltimore. 1972. *In vitro* synthesis of DNA complementary to rabbit reticulocyte 10S RNA. *Nature New Biol.* 235: 163.
- Vogt, V.M. 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. *Eur. J. Biochem.* 33: 192.
- Walker, I.D. 1974. The structure of down feather keratin. Ph.D. thesis, University of Adelaide.

DNA sequences from which double-stranded RNA is transcribed

H. Naora, J. M. Whitelam, R. P. Monckton and R. E. H. Wettenhall

The genetic information encoded in double-stranded RNA prepared from rat liver cells was investigated using an RNA-RNA hybridization technique. Double-stranded RNA prepared from rat liver cells was capable of hybridizing to mRNA from rat liver cells but not to mRNA from mouse sarcoma 180 cells. No hybridization of dsRNA to rRNA (or tRNA) was observed. It is suggested that mRNA consists of a portion of dsRNA at the 5' end, a unique sequence coding for polypeptide in the centre of the molecule and a poly(A) segment at the 3' end.

Double-stranded RNA (dsRNA) is one of several species of RNA normally present in uninfected eukaryotic cells (Monckton and Naora, 1974). There is no conclusive evidence for a definite function of dsRNA in uninfected cells. A knowledge of the nucleotide sequences of DNA from which dsRNA is transcribed should, however, increase our understanding of its function. Recently, dsRNA was shown to hybridize very rapidly to homologous cellular DNA (Harel and Montagnier, 1971; Jelinek and Darnell, 1972). These DNA sequences have not been well defined. Further information concerning these sequences should be obtained from an investigation of the complementarity of dsRNA with other known species of cellular RNA.

In this communication, we describe molecular hybridization experiments between dsRNA and various other species of RNA isolated from rat liver and mouse sarcoma 180 cells. The results suggest that specific sequences similar or identical to those present in the rat liver dsRNA molecules are present in rat liver mRNA molecules.

MATERIALS AND METHODS

Materials

5-³H-orotic acid (22.5 to 29 Ci/mmole), 5,6-³H-uridine (58 Ci/mmole) and 4,5-³H-L-leucine (20 Ci/mmole) were purchased from the Radiochemical Centre, Amersham and New England Nuclear, Boston. Ribonuclease T₁ (from *Aspergillus oryzae*) was obtained from

Worthington Biochemical Corp. and pancreatic ribonuclease A and protease were from Sigma Chemical Co. Oligo(dT)-cellulose was prepared by Dr J.F.B. Mercer of our laboratory. Male Wistar rats (160-190 gm body weight) were used. Sarcoma 180/TG cells (cell stock was donated by Dr I. Marshall, John Curtin School of Medical Research) were collected six days after inoculation into Walter and Eliza outbred strain mice.

Preparation of unlabelled dsRNA from rat liver

dsRNA was prepared and purified from homogenates of rat livers by the procedure described previously (Monckton and Naora, 1974). The dsRNA preparations were free of DNA but were contaminated with fragments of poly(A) which were not covalently linked.

Preparation of mRNAs

For the preparation of ^3H -labelled mRNA from rat liver, ^3H -orotic acid (300 μCi) was injected into the portal vein four hours before decapitation of the rat. Livers obtained were gently homogenized and polysomes were obtained by the method of Schreier and Staehelin (1973). RNA was extracted from polysome pellets according to the method of Perry et al. (1972). mRNA was isolated from polysomal RNA by two methods.

Oligo(dT)-cellulose column chromatography Polysomal RNA dissolved in 10 mM Tris-HCl buffer (pH 7.6), 0.5 M NaCl, was applied to an oligo(dT)-cellulose column previously equilibrated with the above buffer. The column was then washed thoroughly with the same buffer. The retained material was eluted with 10 mM Tris-HCl buffer (pH 7.6), concentrated by Amicon Diaflo ultrafiltration, dialyzed against 10 mM phosphate buffer (pH 7.6), 0.3 M NaCl and stored at -20°C .

Sucrose density gradient fractionation Polysomal RNA was layered on 5-30% sucrose gradients in 10 mM Tris-HCl buffer (pH 7.6), 0.14 M NaCl. Centrifugation was at 25,000 rpm for 18 hours at 4°C in a Spinco SW 27 rotor. RNA sedimenting between 9 to 15S was collected and precipitated with ethanol; the pellet was dissolved in 10 mM phosphate buffer (pH 7.6), 0.3 M NaCl and stored at -20°C .

The specific activities of the mRNA preparation obtained by the two methods were similar, the usual value being approximately 500 cpm per μg RNA.

Sarcoma 180 cells were incubated with ^3H -uridine (16 $\mu\text{Ci}/\text{ml}$) for 60 min at 37°C and subsequently lysed according to the method of Brawerman et al. (1972). Polysomes were prepared by the method of Schreier and Staehelin (1973). Polysomal RNA was extracted and mRNA isolated using an oligo(dT)-cellulose column.

Unlabelled mRNA was prepared from RNA of rat liver or rabbit reticulocyte polysomes by the sucrose density gradient fractionation procedure described above.

It is stressed that all the preparations of mRNA obtained by the above methods were sensitive to potassium hydroxide hydrolysis and insensitive to DNAase.

Preparation of ^3H -labelled tRNA, rRNA and nuclear RNA (nRNA) from rat liver cells

^3H -orotic acid (300 μCi) was injected as described above. The rats were decapitated and livers removed after 60 min for labelling of nRNA and four days for rRNA and tRNA labelling. ^3H -labelled tRNA and rRNA were fractionated from cellular RNA by treatment with 1 M NaCl and purified by sucrose density gradient centrifugation (Naora and Kodaira, 1970). ^3H -labelled nRNA was prepared as described previously (Naora and Kodaira, 1969).

RNA-RNA hybridization

All reactions were carried out at 72°C normally for 24 hours in hybridization solution (10 mM phosphate buffer, pH 7.6, 0.3 M NaCl) by a modification of Schonberg et al. (1971), and Shimotohno and Miura (1973). After annealing, the reaction mixtures were diluted to a final volume of 0.2 ml with 10 mM phosphate buffer, pH 7.6, 0.3 M NaCl and were subjected to treatment with 10 $\mu\text{g}/\text{ml}$ of ribonuclease A and 0.5 $\mu\text{g}/\text{ml}$ of ribonuclease T_1 at 37°C for 30 min. After addition of carrier rat liver RNA, the RNA-RNA hybrids were precipitated with trichloroacetic acid, collected on glass fibre filters and the radioactivity determined (Monckton and Naora, 1974).

Melting profiles of hybrids

After hybridization, the reaction mixture was incubated for 2 min at various temperatures in a thermostatically-controlled water bath. After heating, the surviving hybrids were estimated as described above.

Measurement of the translation of mRNA *in vitro*

Protein synthesis *in vitro* was assessed by a modification of the assay described previously (Wettenhall et al., 1973). The following changes were made: the reaction volume was halved to 25 μl ; the K^+ concentration was reduced to 80 mM; the concentration of ATP was increased to 3 mM; sarcoma 180 'S100' cytosol was substituted for Krebs II ascites cytosol (the sarcoma cytosol was prepared in the same way -- see Mathews and Korner, 1970); ^3H -leucine was used as the labelled amino acid; the incubation was for 60 min at 30°C. In some experiments the Mg^{++} concentration was varied. The

net concentration of 'free' Mg^{++} ions was estimated after assuming that one mole of ATP chelates 0.95 moles of magnesium.

RESULTS

For satisfactory hybridization experiments it was important to use pure preparations of RNA. The procedures used for the isolation of dsRNA and mRNA were specifically designed to achieve the necessary purity. The double-stranded nature of our dsRNA preparation has already been described in detail (Monckton and Naora, 1974).

Characteristics of mRNA preparations

We believe that the 3H -labelled mRNA preparations used here contain mRNA coding for rat liver or sarcoma 180 cellular proteins, for the following reasons:

- (1) The preparations were rapidly labelled, constituted approximately 3% of radioactive polysomal RNA and contained poly(A).
- (2) The poly(A) containing mRNA preparation was heterogeneous in size, but a major portion of the RNA sedimented between 9-15S (Figure 1), characteristic of mRNAs for many cellular proteins. Figure 1 also shows the presence in the preparation of an RNA species larger than 15S. However, the sedimentation profile was not appreciably changed after rechromatography on oligo(dT)-cellulose, a treatment that should have selectively removed any rRNA. The bulk of the 3H -labelled RNA sedimenting faster than 15S was most probably larger species (or aggregates) of messenger-like RNA (Greenberg and Perry, 1971; Taylor and Schimke, 1973; Shafritz, 1974). When preparing mRNA by repeated centrifugation we used a fraction sedimenting between 9 and 15S.
- (3) The mRNA preparations stimulated the incorporation of labelled amino acids into protein (Figure 2, Table 1). There were indications that the stimulations reflected the specific translation of liver mRNA. Firstly, the net magnesium optimum of around 1 mM (Figure 2) is close to the optima for a variety of natural mRNAs undergoing translation with fidelity in cell free systems (Jacobs-Lorena and Baglioni, 1972; Schreier and Staehelin, 1973). Secondly, the stimulation was completely inhibited by aurintricarboxylic acid (Table 1), a specific inhibitor of initiation (Stewart et al., 1971), while the concentration of liver mRNA required to produce maximum stimulation resembled that for reticulocyte mRNA, the latter was translated more efficiently (Table 1).
- (4) Finally, although we have not yet analysed the product of liver mRNA-directed protein synthesis, others have demonstrated the translation of liver-specific proteins, i.e. liver-ferritin and albumin,



Figure 1. Sedimentation analysis of ³H-labelled rat liver mRNA, purified by oligo(dT)-cellulose column chromatography and layered onto a linear sucrose density gradient (5 to 30%) prepared in 10 mM Tris-HCl, pH 7.6 buffer, 0.14 M NaCl. Centrifugation was at 49,000 rpm for 3 hours using a SW 50L rotor in a Spinco L2-65B ultracentrifuge. Fractions were collected by an ISCO density gradient fractionator and radioactivity determined as described previously (Naora and Kodaira, 1969).

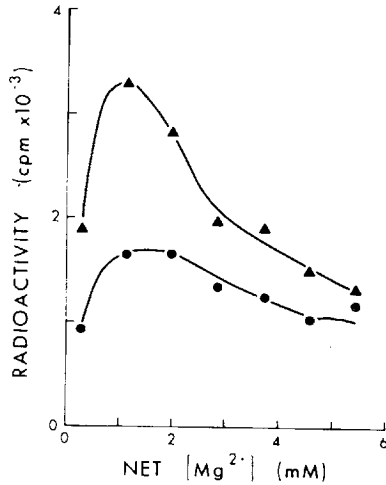


Figure 2. Stimulation of ³H-leucine incorporation by rat liver mRNA as a function of Mg⁺⁺ concentration. ▲—▲, with mRNA (1.7 μg), ●—● without mRNA.

TABLE 1
Stimulatory activity of rat liver mRNA

mRNA	³ H-leucine incorporation (cpm/25 μl assay)
Control (no mRNA)	1125
Rat liver mRNA (1.7 μg)	2060
Rat liver mRNA (1.7 μg) + aurintricarboxylic acid (8×10^{-5} M)	1130
Rabbit globin mRNA (0.9 μg)	7508

from similar preparations of mRNA in analogous cell-free systems (Taylor and Schimke, 1973; Shafritz et al., 1973; Shafritz, 1974).

The possibility of nRNA contamination of the mRNA preparations seems very unlikely for the following reasons:

- (1) The radioactivity incorporated into RNA during the four hour period after injection was located mostly in the cytoplasm, with only a small proportion remaining in the nucleus (results not shown).
- (2) The mild conditions used for homogenizing livers minimized damage to nuclei and, consequently, minimized possible leakage of ³H-labelled HnRNA as well as DNA from nuclei. In fact, less than 0.03% of cellular DNA was present in the postmitochondrial fraction.
- (3) A more conclusive result was obtained when mRNA was prepared from unlabelled livers in the presence of nuclei highly labelled for nRNA, and the final preparation of mRNA contained only traces (0.006%) of radioactivity attributable to nRNA (Table 2). Since nRNA has a much higher specific activity at one hour after administration of the isotope than that at four hours (Muramatsu et al., 1964; Naora and Kodaira, 1969), the possible contamination estimated here would be a maximum.

Hybridization of mRNA and nRNA to dsRNA at a
varying input ratio

Ribonuclease-resistant complexes were formed when purified ³H-labelled rat liver mRNA at a fixed concentration was allowed to anneal with varying amounts of heat-denatured rat liver dsRNA (Figure 3). Only 0.6% of rat liver mRNA became ribonuclease-

TABLE 2
Contamination of mRNA with nRNA

Fraction	Specific activity (cpm/ μ g)		Contamination (% 3 H-RNA as nRNA)
	Normal condition	Radioactive nuclei added	
Polysomal RNA	137	3.9	2.8
mRNA	530	0.033	0.006

3 H-labelled nuclei were prepared from liver cells of rats injected intravenously with 3 H-*orotic acid* 60 min prior to decapitation as described under Materials and Methods. The purified nuclear preparation (5.4×10^5 cpm, 1.5×10^8 nuclei) was mixed with 3.0 g of fresh unlabelled liver where 7.5×10^8 liver cells exist (Naora, 1957). This mixture was homogenized and polysomes, polysomal RNA and finally mRNA were prepared using the normal procedure for labelled mRNA. The specific activities of RNA at each step were determined and compared with those obtained under normal conditions. In this table the specific activities of RNA obtained with labelled nuclei mixed with unlabelled liver cells were expressed in terms of radioactivities estimated at a ratio of 1:1 (numbers of radioactive nuclei : unlabelled cells in the original homogenate with the assumption that the specific activity of RNA observed finally is proportional to the number of radioactive nuclei added. The possible contamination was calculated by using a ratio of the specific activities of these RNAs.

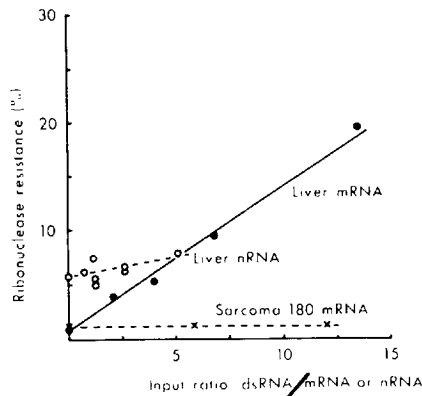


Figure 3. Hybridization of 3 H-labelled RNAs with dsRNA as a function of input ratio. Hybridization was at 4.7 mg/ml/hr for each input ratio: 3 H-labelled RNA, rat liver mRNA (\bullet), mouse sarcoma 180 mRNA (X) and rat liver nRNA (\circ).

resistant after annealing without dsRNA. Therefore, the ribonuclease-resistant RNA formed during annealing with heat-denatured dsRNA represented hybrids between ^3H -labelled mRNA and a complementary strand of heat-denatured dsRNA. The increased ribonuclease-resistance did not result from non-specific aggregation of RNA molecules, which would protect the molecules from enzyme action, because ribonuclease-resistant materials formed during annealing were recovered from the supernatant obtained after centrifugation at 15,000 g for 45 min.

The amount of hybrid formed depended upon the input dsRNA:mRNA ratios. At the highest input ratio studied (13.5:1, w:w), about 20% of the mRNA was hybridized with dsRNA; however, the saturation point had not been reached. When ^3H -labelled mRNA isolated by centrifugation was used to hybridize with dsRNA, the reaction occurred in an essentially similar manner.

Since most dsRNA is of HnRNA origin (Jelinek and Darnell, 1972; Ryskov et al., 1973; Monckton and Naora, 1974), rat liver nRNA should hybridize to dsRNA. This was shown to be the case (Figure 3). Since nRNA used in this experiment was not sheared, a considerable portion (5.8%) of ribonuclease resistant complexes were formed in the absence of dsRNA, presumably as a result of self-annealing. This process would have interfered to a certain extent with the annealing of denatured dsRNA to ^3H -labelled nRNA. In addition rRNA and tRNA are present in nRNA preparations. All these factors prevent the observation of a more pronounced annealing response with HnRNA. The nRNA hybridization to dsRNA observed here would seem to confirm our previous finding of the nuclear origin of dsRNA.

As shown in Figure 3, mRNA from sarcoma 180 cells did not show any significant hybridization to rat liver dsRNA above the background level. DNA-RNA hybridization and competition experiments have demonstrated a significant difference in the RNA population of tumor and normal cells (e.g. Church et al., 1969). The result observed here similarly indicates that the formation of the RNA-RNA hybrid appears specific for homologous mRNA.

Hybridization of mRNA to dsRNA at high RNA concentration

As in the case of DNA-DNA or DNA-RNA hybridization (Britten and Kohne, 1968), RNA-RNA reaction is dependent on both the concentration of complementary nucleotide sequences and the reaction time. Therefore there is a relationship between the product of these variables (analogous to C_{ot} value used in DNA hybridization studies) and the percentage hybridization. In the experiment shown in Figure 4, rat liver mRNA was annealed to liver dsRNA at varying concentrations,

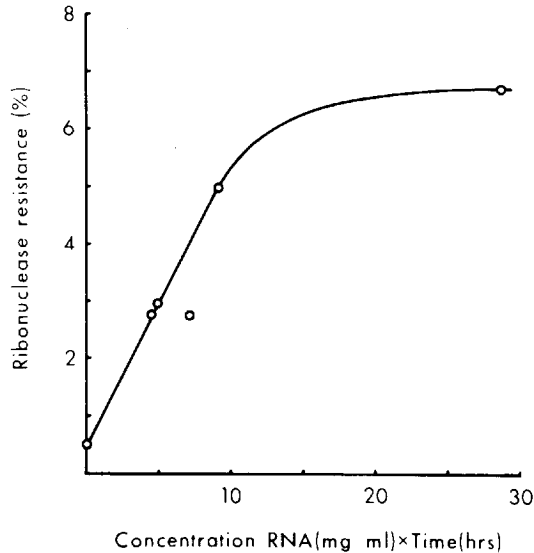


Figure 4. Hybridization of ^3H -labelled rat liver mRNA to rat liver dsRNA at varying concentrations of RNA. Hybridization was carried out at a fixed input ratio (1:1).

but at a fixed input ratio of 1:1. Since less mismatched hybrids are formed at lower input ratio, as will be described, a ratio of 1:1 was chosen in this experiment. High values of mg/ml/hr were normally obtained by increasing the concentration of RNA and not by lengthening the reaction time. Under these conditions 7% of the mRNA added is hybridized with dsRNA. This value would of course be a minimum estimation since complementary sequences in dsRNA are not in excess over all mRNA sequences in question.

Hybridization of rRNA and tRNA to dsRNA

dsRNA failed to hybridize with ^3H -labelled rRNA or tRNA (Figure 5). About 8% of the rRNA and 1.3% of the tRNA remained ribonuclease-resistant after annealing without dsRNA (Figure 5). Similar ribonuclease-resistance in rRNA has been observed by Holmes and Bonner (1974). The presence of double-stranded structures in these species of RNA may prevent molecular hybridization to dsRNA in the solution. Rat liver nRNA containing double-stranded regions hybridized to rat liver dsRNA to a certain extent. This does not account for the failure to obtain hybrid formation, in the case of rRNA and tRNA.

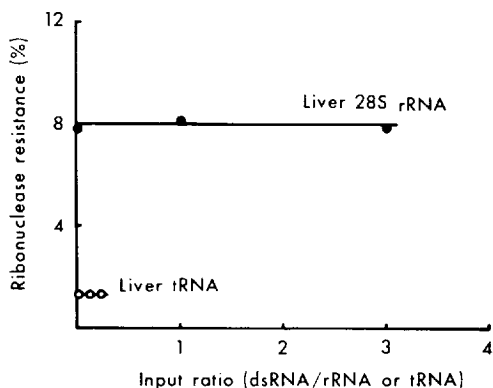


Figure 5. Hybridization of ^3H -labelled rRNA or ^3H -labelled tRNA to dsRNA at varying input ratios. Hybridization was carried out at 4.7 mg/ml/hr but at varying input (dsRNA: rRNA or tRNA) ratios: ^3H -labelled 28S rRNA (●) and tRNA (○).

Melting profiles of hybrids

In order to investigate the extent of mismatching in the hybrids the melting profiles of hybrids formed at varying input ratios of rat liver dsRNA to rat liver mRNA were examined. The result is presented in Figure 6, demonstrating that the thermal stability of the hybrids depended on the input ratio at which the hybrids were obtained. The T_m of rat liver dsRNA at high salt concentration was observed to be 94°C (Monckton and Naora, 1974). The hybrid obtained at an input ratio of 3.3:1 has a broader melting profile with T_m about 26° lower than dsRNA. On the other hand the hybrid at an input ratio of 1:1 melted sharply with a T_m about 6° below that of dsRNA, indicating a reasonably good fidelity of base pairing. It is not known to what extent the reduced T_m was due to the difference of the base composition between dsRNA and mRNA and also due to the shortness of hybrid duplexes (Hayes et al., 1970). Assuming that these factors are not significant, the T_m of the hybrid at low input ratio indicates about 3% mismatching of bases according to the calculations of Greenberg and Perry (1971).

DISCUSSION

We have demonstrated hybridization between preparations of mRNA and dsRNA. The apparent hybrids did not result from non-specific aggregation of RNA nor were they due to hybridization with contaminants of the RNA preparations such as DNA or HnRNA. Moreover, hybridization was specific for preparations of dsRNA and mRNA from the same tissue. We therefore concluded that hybrids resulted from a specific interaction between mRNA and dsRNA.

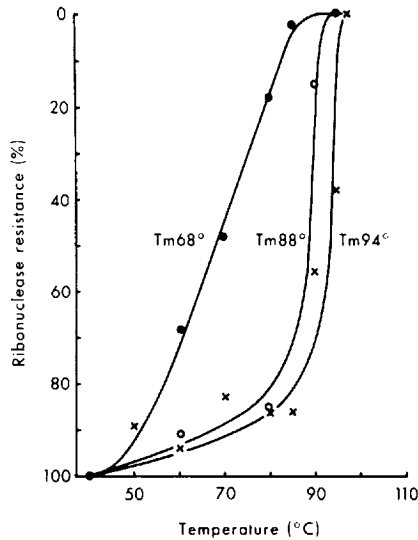
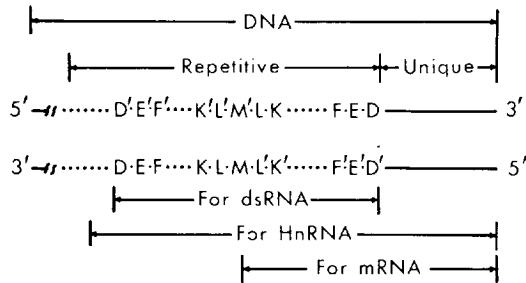


Figure 6. Effect of dsRNA input on the thermal stability of dsRNA-mRNA hybrids. Hybrids were prepared at input ratios (dsRNA:mRNA) of 1:1 (○) and 3.3:1 (●), but at 4.7 mg/ml/hr, and subjected to analysis of thermal dissociation. The melting profile of dsRNA (X) is also shown for comparison. The curves have been normalized to 100% denaturation at 95°C for hybrids and at 97°C for dsRNA. T_m values were 68° and 88° for hybrids at input ratios of 3.3:1 and 1:1, and 94°C for dsRNA.

HnRNA is thought to be a precursor of cytoplasmic mRNA (Darnell et al., 1973). The discovery of dsRNA-mRNA hybrids with low and high thermal stabilities suggests that mRNA contains sequences either similar or identical to those of dsRNA. The presence of these sequences in both nRNA and cytoplasmic mRNA is in accord with Darnell's idea. However, only HnRNA and dsRNA contain ribonuclease resistant double-stranded structures (Jelinek and Darnell, 1972; Monckton and Naora, 1974). In fact, dsRNA is actually derived from the double-stranded or 'hair-pin' structures of HnRNA (Jelinek and Darnell, 1972; Ryskov et al., 1973; Monckton and Naora, unpublished). The absence of double-stranded structures in mRNA is presumably because only a portion of the sequences necessary for double-stranded 'hair-pin' structuring is retained during the processing of HnRNA to mRNA.

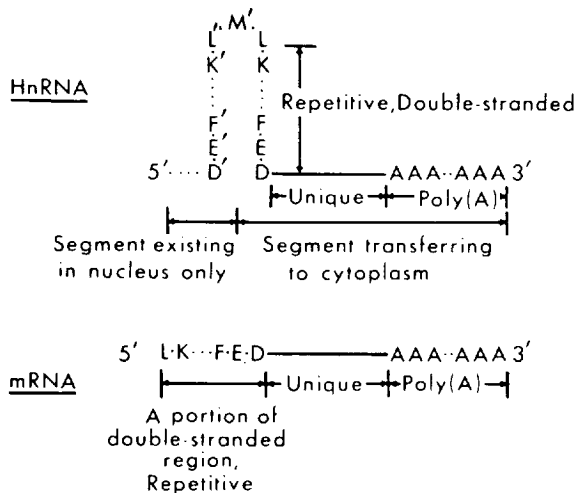
Thus, in addition to a unique sequence and a sequence of poly(A), mRNA would appear to contain a portion from a double-stranded structure (Figures 7 and 8). In rat liver mRNA, the complementary base sequences may constitute 3-12% of the whole mRNA molecule, assuming

that rat liver cytoplasmic mRNA is 500-2,000 nucleotides long and that a sequence 60 nucleotides long, i.e. one strand of dsRNA, is transferred to the cytoplasm as part of mRNA.



DNA sequences from which HnRNA, dsRNA and mRNA are transcribed

Figure 7. A model of DNA sequences from which HnRNA, dsRNA and mRNA are transcribed. A double-stranded DNA molecule which contains repetitive and unique sequences; The 5' and 3' ends are indicated. Letters of the alphabet denote nucleotides which are sequenced in a particular region and are complementary to nucleotides denoted by letters with primes. Arrows show the regions indicated on the DNA molecule. The regions corresponding to dsRNA, HnRNA and mRNA are also indicated.



Schematic HnRNA and mRNA Molecules

Figure 8. Schematic HnRNA and mRNA molecules. HnRNA and mRNA shown here are transcribed from the region indicated in Figure 7. A: Adenosine mononucleotide. See Figure 7 for additional details.

The occurrence of repetitive sequences in HnRNA has been well documented (Molloy et al., 1974; Holmes and Bonner, 1974). Since the majority of dsRNA sequences are transcribed from repetitive DNA sequences (Harel and Montagnier, 1971; Jelinek and Darnell, 1972; Ryskov et al., 1973), the double-stranded structures in HnRNA should also be transcribed from repetitive DNA sequences. Moreover this implies that the segments of cytoplasmic mRNA, which hybridize to dsRNA, are transcribed from repetitive DNA sequences. Since the sequences coding for polypeptides are mainly transcribed from unique DNA sequences, (e.g. Sullivan et al., 1973), mRNA must consist of both repetitive and unique sequences. This view is supported by the recent finding by Dina et al. (1973, 1974), that *Xenopus laevis* embryos contain sequences transcribed from both unique and repetitive DNA sequences. Presumably the sequences in *Xenopus* mRNA transcribed from repetitive DNA sequences are analogous to the sequences in rat liver mRNA derived from dsRNA. However, an apparently contradictory observation is that sea urchin mRNA molecules seem to consist almost exclusively of unique sequence transcripts (Goldberg et al., 1973; McColl and Aronson, 1974). The reason for this discrepancy is not clear at present.

The thermal stability of dsRNA-mRNA hybrids depends upon the input dsRNA:mRNA ratio. This probably results from the fact that these repetitive sequences in mRNA molecules range from groups of almost identical copies to groups with sufficient diversity that, after reassociation, only structures of low stability are formed with dsRNA. In spite of the diversity of the sequences in rat liver mRNA, related sequences do not seem to be present in mRNAs from other species. This was indicated by our observation that liver dsRNA did not hybridize with mouse sarcoma 180 mRNA.

An important question arises as to the location in the mRNA molecule of the sequences derived from dsRNA. Presumably, it is either at the 5' end or between the poly(A) segment at the 3' end and the unique sequence. The former possibility seems more acceptable. In fact, repetitive sequences, 50-60 nucleotides long, were found at the 5' end of *Xenopus* embryo mRNA (Dina et al., 1974), which is about the length of dsRNA (Monckton and Naora, 1974). The function of these sequences at the 5' end of mRNA is not clear. It seems unlikely, however, that they are translated. Eukaryotic mRNAs are known to contain sequences that are not translated (e.g. Howells, 1973).

On the basis of these various observations it is suggested, as shown in Figures 7 and 8, that cytoplasmic mRNA is composed of a sequence corresponding to a portion (one strand) of dsRNA at the 5' end, a unique sequence coding for polypeptide and finally a poly(A) segment at the 3' end. Our findings suggest that the sequence at

the 5' end is a residue of a double-stranded region in the parent HnRNA molecule.

REFERENCES

- Brawerman, G., J. Mendecki and S.Y. Lee. 1972. A procedure for the isolation of mammalian messenger ribonucleic acid. *Biochemistry* 11: 637.
- Britten, R.J. and D.E. Kohne. 1968. Repeated sequences in DNA. *Science* 161: 529.
- Church, R.B., S.W. Luther and B.J. McCarthy. 1969. RNA synthesis in Taper hepatoma and mouse-liver cells. *Biochim. Biophys. Acta* 190: 30.
- Darnell, J.E., W.R. Jelinek and G.R. Molloy. 1973. Biogenesis of mRNA: Genetic regulation in mammalian cells. *Science* 181: 1215.
- Dina, D., M. Crippa and E. Beccari. 1973. Hybridization properties and sequence arrangement in a population of mRNAs. *Nature New Biol.* 242: 101.
- Dina, D., I. Meza and M. Crippa. 1974. Relative positions of the 'repetitive', 'unique' and poly(A) fragments of RNA. *Nature* 248: 486.
- Greenberg, J.R. and R.P. Perry. 1971. Hybridization properties of DNA sequences directing the synthesis of messenger RNA and heterogeneous nuclear RNA. *J. Cell Biol.* 50: 774.
- Goldberg, R.B., G.A. Galau, R.J. Britten and E.H. Davidson. 1973. Non-repetitive DNA sequence representation in sea urchin embryo messenger RNA. *Proc. Nat. Acad. Sci. USA* 70: 3516.
- Harel, L. and L. Montagnier. 1971. Homology of double-stranded RNA from rat liver cells with the cellular genome. *Nature New Biol.* 229: 106.
- Hayes, F.N., E.H. Lilly, R.L. Ratliff, D.A. Smith and D.L. Williams. 1970. Thermal transitions in mixtures of polydeoxyribonucleotides. *Biopolymers* 9: 1105.
- Holmes, D.S. and J. Bonner. 1974. Sequence composition of rat nuclear deoxyribonucleic acid and high molecular weight nuclear ribonucleic acid. *Biochemistry* 13: 841.
- Howells, A.J. 1973. Messenger RNA. *In The Ribonucleic Acids* (P.R. Stewart and D.S. Letham, ed.), p. 59. Springer-Verlag, Berlin.
- Jacobs-Lorena, M. and C. Baglioni. 1972. Characterization of a mouse ascites cell-free system. *Biochemistry* 11: 4970.
- Jelinek, W. and J.E. Darnell. 1972. Double-stranded regions in heterogeneous nuclear RNA from HeLa cells. *Proc. Nat. Acad. Sci. USA* 69: 2537.

- Mathews, M.B. and A. Korner. 1970. Mammalian cell-free protein synthesis directed by viral ribonucleic acid. *Eur. J. Biochem.* 17: 328.
- McCull, R.S. and A.I. Aronson. 1974. Transcription from unique and redundant DNA sequences in sea urchin embryos. *Biochem. Biophys. Res. Comm.* 56: 47.
- Molloy, G.R., W. Jelinek, M. Salditt and J.E. Darnell. 1974. Arrangement of specific oligonucleotides within poly(A) terminated HnRNA molecules. *Cell* 1: 43.
- Monckton, R.P. and H. Naora. 1974. Rat liver double-stranded RNA from nuclei and whole cells: Isolation, purification and characterization. *Biochim. Biophys. Acta* 335: 139.
- Muramatsu, M., J. Hodnett and H. Busch. 1964. Studies on the 'independence' of nucleolar ribonucleic acid synthesis. *Biochim. Biophys. Acta* 91: 592.
- Naora, H. 1957. Microspectrophotometry of cell nuclei stained with the Feulgen reaction. IV. Formation of tetraploid nuclei in rat liver cells during postnatal growth. *J. Biophys. Biochem. Cytol.* 3: 949.
- Naora, H. and K. Kodaira. 1969. Interaction of informational macromolecules with ribosomes. I. Binding of labelled nuclear RNA by rat liver ribosomes. *Biochim. Biophys. Acta* 182: 469.
- _____, _____. 1970. Interaction of informational macromolecules with ribosomes. II. Binding of tissue-specific RNAs by ribosomes. *Biochim. Biophys. Acta* 209: 196.
- Perry, R.P., J. Latorre, D.E. Kelley and J.R. Greenberg. 1972. On the lability of poly(A) sequences during extraction of messenger RNA from polyribosomes. *Biochim. Biophys. Acta* 262: 220.
- Ryskov, A.P., G.F. Saunders, V.R. Farashyan and G.P. Georgiev. 1973. Double-helical regions in nuclear precursor of mRNA (pre-mRNA). *Biochim. Biophys. Acta* 312: 152.
- Schonberg, M., S.C. Silverstein, D.H. Levin and G. Acs. 1971. Asynchronous synthesis of the complementary strands of the reovirus genome. *Proc. Nat. Acad. Sci. USA* 68: 505.
- Schreier, M.H. and T. Staehelin. 1973. Initiation of mammalian protein synthesis: The importance of ribosome and initiation factor quality for the efficiency of *in vitro* systems. *J. Mol. Biol.* 73: 329.
- Shafritz, D.A. 1974. Protein synthesis with messenger ribonucleic acid fractions from membrane-bound and free liver polysomes. *J. Biol. Chem.* 249: 81.
- Shafritz, D.A., J.W. Drysdale and K.J. Isselbacher. 1973. Translation of liver messenger ribonucleic acid in a messenger-dependent reticulocyte cell-free system. *J. Biol. Chem.* 248: 3220.

- Shimotohno, K. and K. Miura. 1973. Transcription of double-stranded RNA in cytoplasmic polyhedrosis virus *in vitro*. *Virology* 53: 283.
- Stewart, M.T., A.P. Grollman and M-T. Huang. 1971. Aurintricarboxylic acid: Inhibitor of initiation of protein synthesis. *Proc. Nat. Acad. Sci. USA* 68: 97.
- Sullivan, D., R. Palacios, J. Stavnezer, J.M. Taylor, A.J. Faras, M.L. Kiely, N.M. Summers, J.M. Bishop and R.T. Schimke. 1973. Synthesis of a deoxyribonucleic acid sequence complementary to ovalbumin messenger ribonucleic acid and quantification of ovalbumin genes. *J. Biol. Chem.* 248: 7530.
- Taylor, J.M. and R.T. Schimke. 1973. Synthesis of rat liver albumin in a rabbit reticulocyte cell-free protein-synthesizing system. *J. Biol. Chem.* 248: 7661.
- Wettenhall, R.E.H., I.G. Wool and C.C. Sherton. 1973. Two physically and functionally distinct forms of eukaryotic 40S ribosomal subunits. *Biochemistry* 12: 2403.

Evidence for mRNA of higher plants

T. J. V. Higgins and P. B. Goodwin

The messenger RNA for a specific protein of higher plants has not yet been isolated; however, there is considerable evidence that such an RNA species does exist in plant cells and that protein synthesis proceeds essentially as in mammalian tissue. The occurrence of poly(A) sequences in plant polysomal RNA provides the opportunity of isolating and studying the properties of mRNA and facilitates a detailed analysis of the regulatory events in protein synthesis.

There is a considerable amount of evidence that higher plants contain RNA with template activity for protein synthesis. The data have been collected over a number of years and many of the results have paralleled those obtained using bacterial and mammalian cells. It is likely therefore that portions of the models of transcription and translation put forward for mammalian cells also apply to plant cells. There are numerous demonstrations of mRNA for specific proteins in mammals and birds but so far a specific plant mRNA has not been isolated. A purified single species of mRNA would allow a detailed study of transcription and factors involved in gene activation. It would also enable us to study protein synthesis and to examine some of the possible regulatory events.

Two recent achievements make the above proposals experimentally feasible. One is the finding that plant RNA contains poly(A) sequences (Manahan et al., 1973; Van der Walle, 1973; Higgins et al., 1973; Kulikowski and Mascarenhas, 1973; Key and Silflow, 1973; Jacobsen and Zwar, 1974a), and that some of this RNA is associated with the polysomes (Manahan et al., 1973; Higgins et al., 1973). Thus it has been assumed that at least some of the presumptive mRNA of plants contain poly(A) sequences and is therefore amenable to purification by affinity chromatography. The second useful procedure is the standardization of the cell-free protein synthesizing system from wheat germ as developed by Marcus and Feeley (1964) and modified by Roberts and Patterson (1973).

It now seems likely that rapid progress will be made in elucidating the steps of transcription and translation in plants. Already there is considerable evidence for the existence of mRNA.

EVIDENCE OF mRNA IN HIGHER PLANTS

Base sequence similar to plant DNA Data on mRNA of bacteria indicated a G + C content similar to the DNA and unlike that of rRNA (transcribed from a very small portion of the DNA). Plants were also shown to possess a similar type of RNA by using radioactive precursors for short labeling periods followed by fractionation on cellulose acetate strips or methylated albumin Keiselguhr (MAK) columns (Loening, 1962; Hotta and Stern, 1963; Key and Ingle, 1964; Chroboczek and Cherry, 1966; Tester and Dure, 1967; Johri and Varner, 1970). In many cases there was no fractionation of the RNA so that nuclear heterogeneous RNA [a small portion of which is probably precursor mRNA by analogy with results from mammalian cells (Ruiz-Carillo et al., 1973)] and transcribed sequences of reiterated DNA might greatly contribute to the result, masking the synthesis of mRNA. However, in a number of cases this postulated mRNA was found to be associated with the microsomes or polysomes (Loening, 1965; Lin et al., 1966; Chroboczek and Cherry, 1966; Johri and Varner, 1970) thus indicating that some of this RNA was probably messenger. Recently there have been suggestions that some of the DNA-like RNA occurs in the nucleus only (Penon, 1972; Key et al., 1972).

High adenylic acid levels characterized DNA-like RNA and subsequent studies using MAK columns showed that two populations of AMP-rich RNA exist in plant cells. It has been suggested that both are associated with polysomes (Johri and Varner, 1970) but Key et al. (1972), using intensive purification procedures, indicate that one fraction is probably nuclear (40% AMP) and the other is polysomal (30% AMP). Penon (1972) reaches similar conclusions but the localization of the two high AMP RNA fractions are reversed. Hormone effects on this presumptive mRNA indicated a physiological role (Key and Ingle, 1964; Tester and Dure, 1967) and suggest that its continued synthesis is required for growth.

Rapidly labeled with radioactive precursors Short term exposure of plant tissues to radioactive phosphate or nucleosides and the subsequent fractionation of the labeled RNA showed that there was a fraction analogous to bacterial mRNA (Hayashi and Spiegelman, 1961) which did not 'chase' into the other species of RNA. It was polydisperse in size and of high specific activity (Loening, 1962; Key and Ingle, 1964; Cherry and van Huystee, 1965; Hemleben-Vielhaben, 1966; Galling and Hemleben-Vielhaben, 1967; Gressel and Galun, 1967; Yoshida et al., 1967; Johri and Varner, 1970; Key et al., 1972). It is also likely that nuclear contamination contributed to some of these data but rapidly labeled RNA has been found associated with polysomes (Lin et al., 1966; Higgins et al., 1973). It also occurs in chloroplasts (Spencer and Whitfeld, 1967).

There is evidence that rapidly labeled RNA can occur as a messenger ribonucleoprotein particle, free in the cytoplasm or associated with polysomes (Ajtkhozhin and Akhanov, 1974).

Bacterial contamination may have contributed to the results in some cases (Lonberg-Holm, 1967; Hock, 1967). However, it appears that this criticism rarely applied (van Huystee and Cherry, 1966). In sterile cultures of *Euglena gracilis* rapidly labeled RNA was synthesized and shown to have template activity (Avadhani and Buetow, 1972).

High sensitivity to RNase mRNA of mammalian cells is sensitive to low levels of RNase while rRNA is unaffected (Marbaix and Burny, 1964). The presumptive messenger RNA of plants is also very sensitive (Bayley, 1964; Clark et al., 1964; Lin et al., 1966; Watts and Matthias, 1967; Mascarenhas and Bell, 1969; Johri and Varner, 1970; Davies et al., 1972). RNA isolated by affinity chromatography was sensitive to low RNase levels (10^{-8} $\mu\text{g/ml}$) (Higgins, 1974).

Such marked sensitivity is probably due to the nature of the secondary structure of the mRNA molecule.

Heterogeneity of size It is to be expected that mRNA would display a range of sizes considering the number of proteins with differing molecular weights which are potentially coded in a plant cell. This assumption is borne out by the results obtained with the RNA associated with microsomes or polysomes (Figure 1) (Loening, 1962; Ingle et al., 1965; Johri and Varner, 1970; Key et al., 1972; Penon, 1972; Higgins et al., 1973) and also with chloroplast RNA (Spencer and Whitfeld, 1967).

Selective action of antibiotics The synthesis of different species of plant RNA can be inhibited to varying extents by the selective use of antibiotics. Actinomycin D inhibits all species (Key and Ingle, 1964; Chroboczek and Cherry, 1966) while 5-fluorouracil appears to inhibit the synthesis of tRNA and rRNA with little effect on mRNA (Key and Ingle, 1964). These results have been confirmed many times and the synthesis of RNA isolated on oligo-(dT)-cellulose (Figure 1) or poly(U)-cellulose (Jacobsen and Zwar, 1974a) was relatively unaffected by 5-FU.

α -amanitin, on the other hand, inhibits the synthesis of AMP-rich RNA (Seitz and Seitz, 1971) and poly(A)-containing RNA (Higgins et al., 1973) while having little effect on rRNA. This is similar to the situation in mammalian cells where mRNA synthesis is inhibited (Stirpe and Fiume, 1967).

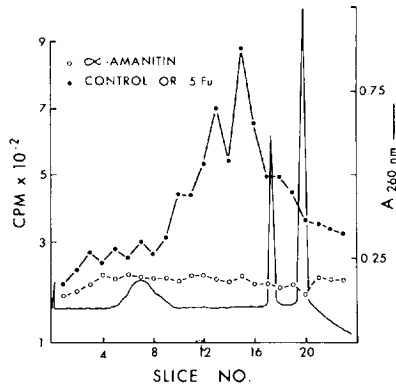


Figure 1. Poly(A)-containing RNA from plant polysomes. The effect of 5-fluorouracil and α -amanitin on the synthesis of poly(A) RNA. For details of procedures see Higgins et al., 1973.

Cordycepin (3'-deoxyadenosine) inhibits the addition of poly(A) to mRNA in mammalian cells (Darnell et al., 1973) and thus probably interferes with processing and transport to the cytoplasm. It has no influence on other species of RNA. Studies with cordycepin in plant cells have been limited but it has been shown to inhibit both rRNA and mRNA to about the same extent in barley aleurone cells (Jacobsen and Zwar, 1974b) and in mung bean roots (Higgins, unpublished).

Polysomes Further evidence for the existence of mRNA was provided by the isolation of polysomes. These aggregates of ribosomes are visible in the electron microscope (Lin et al., 1966) and have rapidly labeled DNA-like RNA associated with them (Loening, 1965; Lin et al., 1966; Johri and Varner, 1970; Key et al., 1972). Unlike monosomes they promote *in vitro* amino acid incorporation into acid precipitable material (Marcus and Feeley, 1964; Lin et al., 1966; Leaver and Key, 1967; Chen et al., 1968).

Plant ribosomes aggregate to form polysomes under specified conditions if supplied with a natural mRNA such as the viral template of tobacco mosaic virus (Marcus et al., 1968) demonstrating the potential of plants to utilize mRNA.

The synthesis of the small subunit of ribulose 1,5-diphosphate carboxylase has been demonstrated by immunoprecipitation from cytoplasmic ribosomes of bean leaves using an *in vitro* system (Gray and Kekwick, 1973). There is also strong evidence that the large subunit of the same enzyme is synthesized on RNase-sensitive components of an *in vitro* system from chloroplasts of spinach (Bottomley et al., 1974).

Poly(A) sequences in plant RNA Sequences of poly-adenylic acid have been detected in plant polysomal RNA (Figure 2 and Manahan et al., 1973), total RNA (Van der Walle, 1973; Key and Silflow, 1973; Jacobsen and Zwar, 1974a) and in nuclear heterodisperse RNA (Kulikowski and Mascarenhas, 1974). These findings suggest that at least some of the mRNAs of all eucaryotic cells contain poly(A).

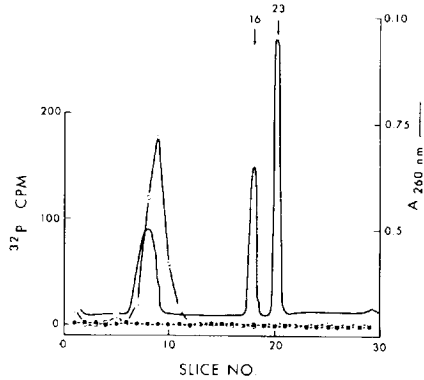


Figure 2. Poly(A) segment resistant to ribonuclease A and T_1 . The distribution of ^{32}P labeled poly(A) sequence on polyacrylamide gels. For details see Higgins et al., 1973.

● ——— ● Pancreatic RNase + T_1 + T_2 , ○ ——— ○ Pancreatic RNase + T_1 .

The high specific activity, rapidly labeled poly(A)-containing RNA of plants is associated with polysomes (Higgins et al., 1973) and its synthesis is inhibited by α -amanitin (Figure 1) but not by 5-fluorouracil (Figure 1 and Jacobsen and Zwar, 1974b).

Gibberellic acid induced *de novo* synthesis of α -amylase in barley aleurone cells apparently requires the selective synthesis of poly(A)-containing RNA (Jacobsen and Zwar, 1974a). This data provides the first strong evidence for template RNA with a functional role and its control by a plant hormone.

The exact size of the poly(A) tract in plants has not been worked out, in the case of polysomal RNA using oligo(dT) it appears to be 5-6S (Figure 2); however, oligo(dT) will not efficiently bind segments shorter than about 50 nucleotides and thus segments of say 20 nucleotides would be lost. Poly(U)-cellulose chromatography indicates that in total RNA there may be two populations of poly(A), one about 7S and the other less than 4S. Whether the short sequence is associated with non-polysomal RNA (e.g. nuclear, Nakazato et al., 1974) or is the

result of the superior binding capacity of poly(U)-cellulose remains to be elucidated.

In vitro protein synthesis Cell-free extracts of plants contain all the components necessary to synthesize protein when supplied with a known messenger RNA such as haemoglobin (Efron and Marcus, 1973a; Roberts and Patterson, 1973); keratin (Kemp et al., 1974) or the viral mRNA of tobacco satellite necrosis virus (Klein et al., 1972); brome mosaic virus (Shih and Kaesberg, 1973) and tobacco mosaic virus (TMV) (Roberts and Patterson, 1973; Efron and Marcus, 1973b; Roberts et al., 1973). The product of the wheat germ *in vitro* system has been identified using specific antibody precipitation, tryptic peptide analysis and detergent-polyacrylamide electrophoresis (Figure 3). There is considerable

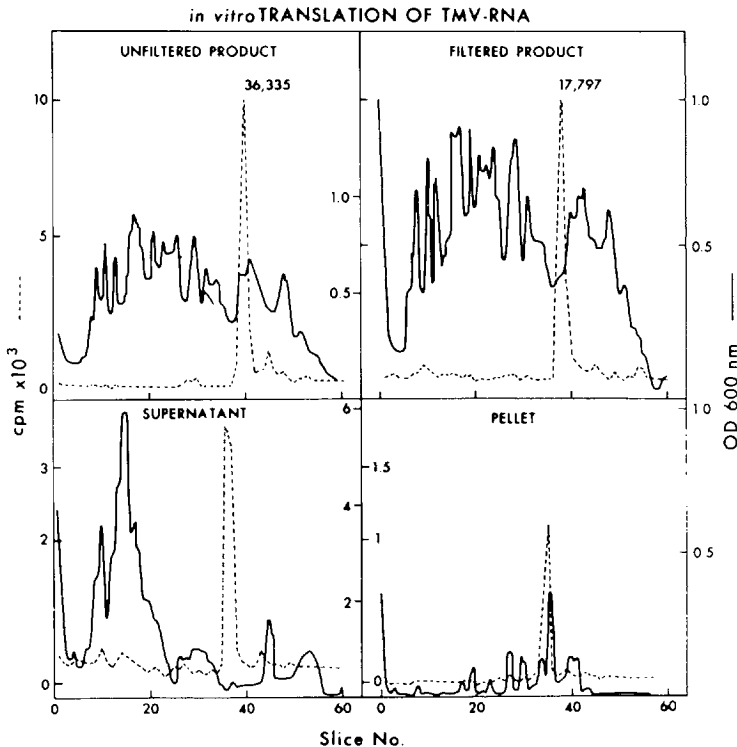


Figure 3. TMV (Bean variety) coat protein synthesized *in vitro*. *In vitro* product of the wheat embryo cell-free system (Roberts and Patterson, 1973) using viral template as messenger. *Top left*, unpurified; *right*, filtered through Millipore to remove proteins still associated with RNA; *bottom left*, antibody-protein complex using authentic coat protein carrier; *right*, counts still remaining in the supernatant after antibody precipitation.

evidence that cell-free plant extracts contain endogenous capacity for protein synthesis (see Mans, 1967) and it has been shown that the total RNA of wheat embryo when added to a homologous *in vitro* system will stimulate incorporation of amino acids (Chen et al., 1968). Fractionation of the RNA verified that there was indeed RNA present with template activity (Weeks and Marcus, 1971; Shultz et al., 1972). This evidence would be more convincing if a heterologous cell-free system had been used as was the case in the translation of fractionated RNA from dark-grown *Euglena gracilis* which stimulated an *in vitro* system from *E. coli* to produce a number of polypeptides as defined by polyacrylamide gel electrophoresis (Avadhani and Buetow, 1972). It is not known, however, if these proteins are normal constituents of *E. gracilis*.

In conclusion, there is considerable evidence to suggest that plant cells contain mRNA and that they function essentially as other eucaryotic cells. Definitive proof for mRNA will only come when a defined protein is synthesized in a heterologous cell-free system.

REFERENCES

- Ajtkhozhin, M.A. and U.A. Akhanov. 1974. Release of mRNA-particles of the informosome type from polyribosomes of higher plant embryos. *Fed. Europ. Biochem. Soc. Letters* 41: 275.
- Avadhani, N.G. and D.E. Buetow. 1972. Studies on messenger RNA from *Euglena gracilis*. 1. Techniques of saline fractionation and characterization of messenger RNAs from dividing cells. *Biochim. Biophys. Acta* 262: 488.
- Bayley, S.T. 1964. Physical studies on ribosomes from pea seedlings. *J. Mol. Biol.* 8: 231.
- Bottomley, W., D. Spencer and P. Whitfeld. 1974. Protein synthesis in isolated spinach chloroplasts: comparison of light-driven and ATP-driven synthesis. *Arch. Biochem. Biophys.* In press.
- Chen, D., S. Sarid and E. Katchalski. 1968. Studies on the nature of messenger RNA in germinating wheat embryos. *Proc. Nat. Acad. Sci. USA* 60: 902.
- Cherry, J.H. and R. van Huystee. 1965. Effects of 5-fluorouracil on photoperiodic induction and nucleic acid metabolism of *Xanthium*. *Plant Physiol.* 40: 987.
- Chroboczek, H. and J.H. Cherry. 1966. Characterization of nucleic acids in peanut cotyledons. *J. Mol. Biol.* 19: 28.
- Clark, M.F., R.E.F. Mathews and R.K. Ralph. 1964. Ribosomes and polyribosomes in *Brassica pekinensis*. *Biochim. Biophys. Acta* 91: 289.

- Darnell, J.E., W.R. Jelinek and G.R. Molloy. 1973. Biogenesis of mRNA : genetic regulation in mammalian cells. *Science* 181: 1215.
- Davies, E., B.A. Larkins and R.H. Knight. 1972. Polyribosomes from peas : an improved method for their isolation in the absence of ribonuclease inhibitors. *Plant Physiol.* 50: 581.
- Efron, D. and A. Marcus. 1973a. Efficient synthesis of rabbit globin in a cell-free system from wheat embryo. *Fed. Europ. Biochem. Soc. Letters* 33: 23.
- _____, _____. 1973b. Translation of TMV-RNA in a cell-free wheat embryo system. *Virology* 53: 343.
- Galling, G. and V. Hemleben-Vielhaben. 1967. Sur lokalization schnell markierter RNA in *Chlorella* und *Phaseolus*. *Arch. Mikrobiol.* 56: 205.
- Gray, J.C. and R.G.O. Kekwick. 1973. Synthesis of the small subunit of ribulose 1,5-diphosphate carboxylase on cytoplasmic ribosomes from greening bean leaves. *Fed. Europ. Biochem. Soc. Letters* 38: 67.
- Gressel, J. and E. Galun. 1967. Morphogenesis in *Trichoderma*: photoinduction and RNA. *Dev. Biol.* 15: 575.
- Hayashi, M. and S. Spiegelman. 1961. The selective synthesis of informational RNA in bacteria. *Proc. Nat. Acad. Sci. USA* 47: 1564.
- Hemleben-Vielhaben, V. 1966. Characterization of rapidly labeled nucleic acids in tissues of plant seedlings. *Zeit. Naturforsch.* 21b: 983.
- Higgins, T.J.V. 1974. Incorporation of ^{32}P into RNA and proteins of the microsomal fraction of mung bean roots : Effects of ABA during the induction of nitrate reductase. In Proceedings of Eighth International Conference on Plant Growth Substances (S. Tamura ed.). In press.
- Higgins, T.J.V., J.F.B. Mercer and P.B. Goodwin. 1973. Poly(A) sequences in plant polysomal RNA. *Nature New Biology* 246: 68.
- Hock, B. 1967. Nature of a rapidly labeled RNA fraction described in higher plant systems. *Plant Physiol.* 42: 1149.
- Hotta, Y. and H. Stern. 1963. Synthesis of messenger-like ribonucleic acid and protein during meiosis in isolated cells of *Trillium erectum*. *J. Cell Biol.* 19: 45.
- Ingle, J., J.L. Key and R.E. Holm. 1965. Demonstration and characterization of a DNA-like RNA in excised plant tissue. *J. Mol. Biol.* 11: 730.
- Jacobsen, J.V. and J.A. Zwar. 1974a. Stimulation of synthesis of RNA containing poly(A) by gibberellic acid in barley aleurone tissue. *Proc. Nat. Acad. Sci. USA*. In press.

- _____, _____. 1974b. Gibberellic acid and ribonucleic acid synthesis in barley aleurone layers : metabolism of ribosomal and transfer RNA and of RNA containing polyadenylic acid sequences. *Aust. J. Plant Physiol.* In press.
- Johri, M.M. and J.E. Varner. 1970. Characterization of rapidly labeled ribonucleic acid from dwarf peas. *Plant Physiol.* 45: 348.
- Kemp, D.J., M. Schwinghamer and G.E. Rogers. 1974. Translation of pure feather keratin messenger RNA in a wheat embryo cell-free system. *Mol. Biol. Reports.* In press.
- Key, J.L. and J. Ingle. 1964. Requirement for the synthesis of DNA-like RNA for growth of excised plant tissue. *Proc. Nat. Acad. Sci. USA* 52: 1382.
- Key, J.L., C.J. Leaver, J.R. Cowles and J.M. Anderson. 1972. Characterization of short time labeled adenosine monophosphate-rich ribonucleic acids of soybean. *Plant Physiol.* 49: 783.
- Key, J.L. and C. Silflow. 1973. Polyadenylate sequences in AMP-rich RNA of soybean. *Plant Physiol.* 51 Suppl.: 38.
- Klein, W.H., C. Nolan, J.M. Lazar and J.M. Clark. 1972. Translation of satellite tobacco necrosis virus ribonucleic acid. 1. Characterization of *in vitro* procaryotic and eucaryotic translation products. *Biochemistry* 11: 2009.
- Kulikowski, R.R. and J.P. Mascarenhas. 1974. Evidence for the presence of heterodisperse nuclear RNA in higher plants. *Biochim. Biophys. Acta.* In press.
- Leaver, C.J. and J.L. Key. 1967. Polyribosome formation and RNA synthesis during aging of carrot-root tissue. *Proc. Nat. Acad. Sci. USA* 57: 1338.
- Lin, C.Y., J.L. Key and C.E. Bracker. 1966. Association of D-RNA with polyribosomes in the soybean root. *Plant Physiol.* 41: 976.
- Loening, U.E. 1962. Messenger ribonucleic acid in pea seedlings. *Nature* 195: 467.
- _____. 1965. Synthesis of messenger ribonucleic acid in excised pea-seedling root segments. *Biochem. J.* 97: 125.
- Lonberg-Holm, K.K. 1967. Nucleic acid synthesis in seedlings. *Nature* 213: 454.
- Manahan, C.O., A.A. App and C.C. Still. 1973. The presence of polyadenylate sequences in ribonucleic acid of a higher plant. *Biochem. Biophys. Res. Comm.* 53: 588.
- Mans, R.J. 1967. Protein synthesis in higher plants. *Ann. Rev. Plant Physiol.* 18: 127.
- Marbaix, G. and A. Burny. 1964. Separation of the messenger RNA of reticulocyte polyribosomes. *Biochem. Biophys. Res. Comm.* 16: 522.

- Marcus, A., B. Luginbill and J. Feeley. 1968. Polysome formation with tobacco mosaic virus RNA. *Proc. Nat. Acad. Sci. USA* 59: 1243.
- Marcus, A. and J. Feeley. 1964. Activation of protein synthesis in the inhibition phase of seed germination. *Proc. Nat. Acad. Sci. USA* 51: 1075.
- Mascarenhas, J.P. and E. Bell. 1969. Protein synthesis during germination of pollen : studies on polyribosome formation. *Biochim. Biophys. Acta* 179: 199.
- Nakazato, H., M. Edmonds and D.W. Kopp. 1974. Differential metabolism of large and small poly(A) sequences in the heterogeneous nuclear RNA of HeLa cells. *Proc. Nat. Acad. Sci. USA* 71: 200.
- Penon, P. 1972. Ribonucleic acid synthesis and hormone action in lentil roots. *Symp. Biol. Hung.* 13: 325.
- Roberts, B.E. and B.M. Patterson. 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proc. Nat. Acad. Sci. USA* 70: 2330.
- Roberts, B.E., M.B. Matthews and C.J. Bruton. 1973. Tobacco mosaic virus RNA directs the synthesis of a coat protein peptide in a cell-free system from wheat. *J. Mol. Biol.* 80: 733.
- Ruiz-Carillo, A., M. Beato, G. Schutz, P. Feigelson and V.G. Allfrey. 1973. Cell-free translation of the globin message within polydisperse high-molecular weight ribonucleic acid of avian erythrocytes. *Proc. Nat. Acad. Sci. USA* 70: 3641.
- Schultz, G.A., D. Chen and E. Katchalski. 1972. Localization of a messenger RNA in a ribosomal fraction from ungerminated wheat embryos. *J. Mol. Biol.* 66: 379.
- Seitz, U. and U. Seitz. 1971. Selektive Hemmung der Synthese der AMP-reichen RNS durch α -Amanitin in Zellen höherer Pflanzen. *Planta* 97: 224.
- Shih, D.S. and P. Kaesberg. 1973. Translation of Brome Mosaic Viral ribonucleic acid in a cell-free system derived from wheat embryo. *Proc. Nat. Acad. Sci. USA* 70: 1799.
- Spencer, D. and P. Whitfeld. 1967. Ribonucleic acid synthesizing activity of spinach chloroplasts and nuclei. *Arch. Biochem. Biophys.* 121: 336.
- Stirpe, F. and L. Fiume. 1967. Studies on the pathogenesis of liver necrosis by α -amanitin : effect of α -amanitin on ribonucleic acid synthesis and on ribonucleic acid polymerase in mouse liver nuclei. *Biochem. J.* 105: 779.

- Tester, C.F. and L.S. Dure. 1967. Nucleic acid synthesis during the hormone-stimulated growth of excised oat coleoptiles. *Biochemistry* 6: 2532.
- Van der Walle, C. 1973. Polyadenylic sequences in plant RNA. *Fed. Europ. Biochem. Soc. Letters* 34: 31.
- van Huystee, R.B. and J.H. Cherry. 1966. Hybridization of messenger RNA with DNA from plants. *Biochem. Biophys. Res. Comm.* 23: 835.
- Watts, R.L. and A.P. Matthias. 1967. The use of bentonite in the isolation of plant polyribosomes. *Biochim. Biophys. Acta* 145: 828.
- Weeks, W.P. and A. Marcus. 1971. Preformed messenger of quiescent wheat embryos. *Biochim. Biophys. Acta* 232: 671.
- Yoshida, K., K. Umerma, K. Yoshinaga and Y. Oota. 1967. Specific RNA from photoperiodically induced cotyledons of *Pharbitis nil*. *Plant and Cell Physiol.* 8: 97.

A model of chromatin structure

Roger D. Kornberg

Histones prepared by mild methods are found as two oligomers, an $\alpha_2\beta_2$ tetramer of F2A1 and F3 and a polymer of F2A2 and F2B. This leads to a model of chromatin structure based on a repeating unit of two each of F2A1, F3, F2A2, and F2B and 200 base pairs of DNA. The concentration dependence of the X-ray pattern of chromatin suggests that the repeating units form a flexibly jointed chain.

I wish to discuss the problem of how DNA and histones are arranged in chromatin. I shall begin with a brief account of evidence suggesting a model of this arrangement (Kornberg and Thomas, 1974; Kornberg, 1974) and then consider some features of the model in more detail.

REPEATING FOLD OF DNA

Two observations have been taken to suggest a repeating fold of the DNA in chromatin. The first dates to early work of Wilkins and his colleagues who obtained X-ray patterns from chromatin showing relatively sharp bands (Wilkins, 1956; Wilkins et al., 1959). Such bands are evidence of repeats of structure, and have been attributed (Pardon and Wilkins, 1972) to a pattern of folding of DNA. The bands in X-ray patterns of chromatin may be due partly to scattering from DNA and partly to scattering from protein (for example, some bands coming from DNA and others from protein), or due entirely to scattering from DNA, or due entirely to scattering from protein. It seems likely that the bands are due in some part to scattering from DNA since chromatin contains nearly equal weights of DNA and protein and the intensity of scattering of X-rays from DNA is about twice as great as from an equal weight of protein.

The second observation suggesting a repeating fold has come from biochemical studies. Cleavage of the DNA in rat liver nuclei by an endogenous nuclease results in fragments which are multiples of the smallest size unit. The occurrence of multiples rather than just the smallest size has been attributed to some cleavage sites being 'sterically blocked by superstructure proteins' (Hewish and Burgoyne, 1973).

These observations raise a number of questions. It may be asked, for example, whether they stem from the same repeating fold: is there a unique pattern of folding of all the DNA in chromatin or does the pattern vary from one part of chromatin to another? It may also be asked what proteins give rise to the repeating fold. Are they the histones? The histones have been shown (Wilkins et al., 1959; Richards and Pardon, 1970) to give rise to the repeating fold revealed in the X-ray patterns. Are all five types of histone involved or only a limited number? Above all, it may be asked what is the pattern of folding of the DNA and what arrangement of proteins gives rise to it.

REPEATING ARRANGEMENTS OF HISTONES

Such questions call for studies of chromosomal proteins, their patterns of association with one another and with DNA. The histones may be fractionated by mild methods (van der Westhuyzen and von Holt, 1971) to yield an F2A1-F3 mixture, an F2A2-F2B mixture, and pure F1. We have found that the F2A1-F3 mixture exists entirely in the form of $\alpha_2\beta_2$ tetramers, that is $(F2A1)_2 (F3)_2$ tetramers (Kornberg and Thomas, 1974). The evidence for this comes from cross-linking and sedimentation experiments. Cross-linking was carried out by treating the F2A1-F3 mixture with dimethyl suberimidate, a bifunctional amino group reagent (Davies and Stark, 1970). The mixture of cross-linked products was then analyzed in a 7.5% SDS-polyacrylamide gel. The results are shown in Figure 1. The various bands in the gel were identified by finding combinations of F2A1 and F3 whose molecular weights, when plotted against the distances of migration of the bands, gave a straight line. All the bands expected for an $(F2A1)_2 (F3)_2$ tetramer are present, but none of the additional bands that would arise from tetramers of other compositions [for example no $(F2A1)_3$ band as would arise from an $(F2A1)_3 F3$ tetramer].

Strictly speaking, the pattern of cross-linked products is consistent with either some mixture of monomers, dimers, etc., or a homogeneous solution of $(F2A1)_2 (F3)_2$ tetramer. The possibility of a mixture was ruled out and the tetramer shown to be homogeneous by sedimentation experiments. We observed (Figure 2) a single sharp sedimentation boundary (sedimentation coefficient 3S) and the behaviour expected of a homogeneous species at sedimentation equilibrium (a linear plot of log. concentration as a function of the square of the distance from the axis of rotation). The sedimentation equilibrium results give a value of 53,900 for the molecular weight (assuming a partial specific volume of 0.72 cm³/g). This is in good agreement with the value of 53,200 expected for the molecular weight of an $(F2A1)_2 (F3)_2$ tetramer.

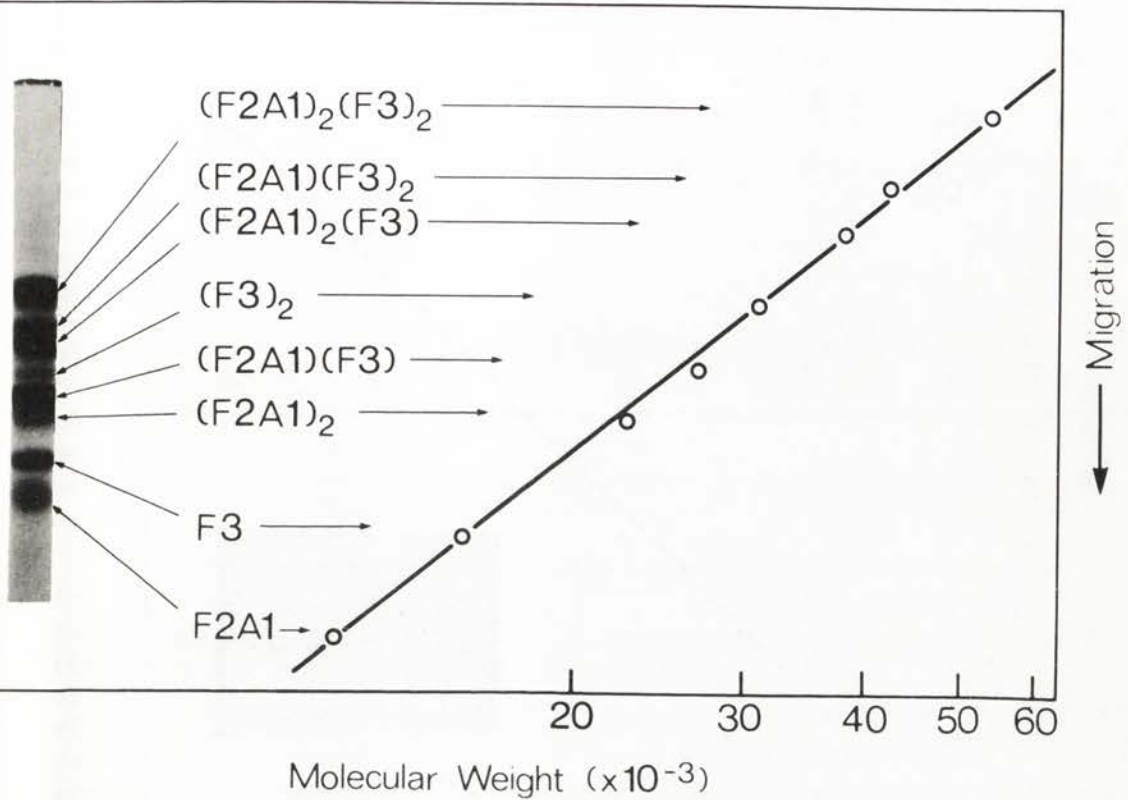


Figure 1. Cross-linking of F2A1-F3 mixture

I shall mention some results of cross-linking experiments on the F2A2-F2B mixture prepared according to van der Westhuyzen and von Holt. We find cross-linked products ranging from monomers to hexamers or heptamers, which we tentatively attribute to the formation of short chains, such as $(-F2A2-F2B-)_n$ or $(-F2A2-F2A2-F2B-F2B-)_n$, by reversible polymerization.

MODEL

It takes no great ingenuity to develop a model of chromatin structure based on the tetramers and the polymer of histones. The model almost suggests itself: globular tetramers at intervals on a chromatin fibre and a polymer running the length. A repeating unit of the structure would be defined by the tetramer and would thus contain two each of F2A1 and F3. Since there are roughly equimolar amounts of the various types of histone in chromatin and one of each type per 100 base pairs of DNA, the repeating unit would also contain two each of F2A2 and F2B and about 200 base pairs

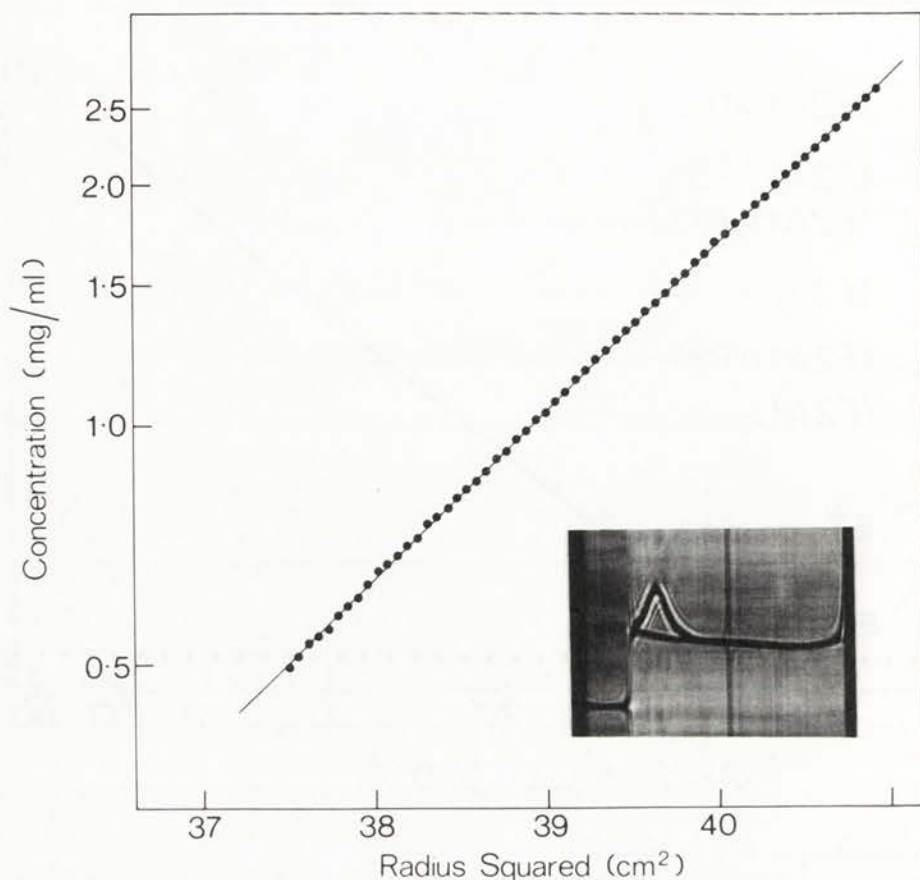


Figure 2. Sedimentation experiments of F2A1-F3 mixture. Schlieren photograph was taken 32 min after start of run. Sedimentation equilibrium measurements were made at a speed of 12,000 rev/min.

of DNA. (In making this use of stoichiometries it is assumed that most of the histones and DNA in chromatin are involved in the repeating structure.)

What are the dimensions of the repeating unit? The longest repeat spacing measured from the bands in X-ray patterns of chromatin is about 100 Å. Work with fibres of chromatin has shown this repeat to be oriented in the fibre direction (Richards and Pardon, 1970). It is natural to associate the spacing with the interval between tetramers, or in other words the length of the repeating unit in the fibre direction. The width of the repeating

unit can also be obtained from X-ray work. Low-angle X-ray scattering from dilute solutions of chromatin fits theoretical curves for rods of about 90 Å diameter (Luzzati and Nicolaieff, 1963; Bram and Ris, 1971). Thus the repeating unit may be about 100 Å in length and 90 Å in diameter.

What is the arrangement of DNA in the unit? Some general features emerge. The first is that the DNA must be tightly coiled or folded. This is an obvious consequence of 200 base pairs or 680 Å length of DNA being contained in a 100 Å by 90 Å unit cell. Another general feature is that the DNA is probably on the outside of the unit. This follows from the globular nature of the histone tetramer and the likelihood of its binding DNA in the same way as many globular proteins bind substrate and effector molecules. Thus much of the 200 base pairs of DNA in a repeating unit would follow some path on the tetramer, with the remainder connecting tetramers along a path defined by the histone polymer. A chromatin fibre would consist of tightly packed DNA and associated protein alternating with more extended DNA and associated protein, rather like beads on a string.

I should like to draw attention to two points of biological interest. The first concerns the DNA being on the outside. This could be of significance in making DNA base sequences available for recognition, for example, by gene 'activator' proteins (Davidson and Britten, 1973). It is of course conceivable that there are sites, either in or between the beads, where base sequences are exposed to particular advantage, for example by strain on the double helix or even, as Crick (1971) has suggested, unpairing of the strands.

The second point of biological interest is that a jointed structure, like beads on string, may be flexible. This was first suggested by X-ray work (see below). The idea also arises from the fact that a chromatin fibre is flexible enough to be extensively coiled or folded. Such coiling or folding must occur, for example, in the bands of polytene chromosomes of *Drosophila*, where the ratio of length of DNA to length of DNA-containing structure is an order of magnitude greater than in a chromatin fibre.

To summarize, there emerges a model of chromatin structure at 50-100 Å resolution, having some features of biological interest. The main points are:

- (1) Chromatin structure is based on a repeating unit of two each of the four main types of histone and about 200 base pairs of DNA.
- (2) A chromatin fibre consists of many such units forming a flexibly jointed chain.
- (3) The unit is about 100 Å in the fibre direction and about 90 Å in diameter.

- (4) The DNA is on the outside of the unit, and the histone on the inside.
- (5) The histones are arranged as an $(F2A1)_2 (F3)_2$ tetramer and F2A2-F2B polymer.
- (6) The structure is ubiquitous: it occurs in all organisms having histones and encompasses almost all the histone and DNA.

SUPPORTING EVIDENCE

The model, based on the association behaviour of the histones, is compatible with a large amount of other data on chromatin structure. I shall discuss, in particular, data on the DNA content of the repeating unit, the location of the DNA in the unit, and the nature of the links in a chain of units.

200 base pairs of DNA Hewish and Burgoyne (1973) have reported that an endonuclease cleaves the DNA in rat liver nuclei into multiples of a smallest unit size. It was obvious, in view of the model described above, to ask if the unit size might not be 200 base pairs. Hewish (personal communication) has estimated a length of 220 base pairs, based on velocity sedimentation analysis. The agreement was encouraging, but there was room for doubt due to the approximate nature of a size measured by sedimentation analysis [it was worrying that Sahasrabudde and van Holde (1974) estimated the size at about 100 base pairs by sedimentation analysis]. The doubt was removed when Noll (1974) obtained a value of 205 base pairs by gel electrophoretic analysis using fragments of bacteriophage ϕ X174 DNA of known sequence as size markers.

DNA on the outside Clark and Felsenfeld (1971) have reported elegant experiments on the location of the DNA in chromatin. They describe results on nuclease digestion of chromatin in which about 50% of the DNA can be converted to acid-soluble form, and results on polylysine binding in which saturation appears to be reached at a lysine-to-nucleotide ratio of 1.0 (independent of the degree of polymerization of the polylysine in the range 50 to 670). Clark and Felsenfeld conclude that about half the DNA is in long stretches completely free of protein, while the other half is 'covered' by protein. The model described above (with DNA on the outside of the structure and protein on the inside) suggests a quite different interpretation, namely that one side of the double helix is exposed to solution and the other side associated with protein (for details of the interpretation of the nuclease digestion experiments see Kornberg, 1974, note 8).

The histones are often said to 'cover' the DNA in chromatin. It would seem, in view of the model, that to speak of the DNA covering the histones would be nearer the truth.

Beads on a string: a flexibly jointed chain The 100 Å band in X-ray patterns of chromatin is attributed, in the context of the model, to the spacing apart of tetramers or length of the repeating unit in the fibre direction (see above). The 100 Å band is enhanced relative to background scatter as the concentration of chromatin is raised (Figure 3). A possible interpretation is as follows. At low concentrations a chromatin fibre follows an irregular path so there are few straight runs of repeating units and little intensity in the 100 Å band. Raising the concentration leads to closer packing and straightening out of fibres, more straight runs of repeating units and greater intensity in the 100 Å band. Thus the concentration dependence of the X-ray pattern of chromatin is compatible with the repeating units forming a flexibly jointed chain.

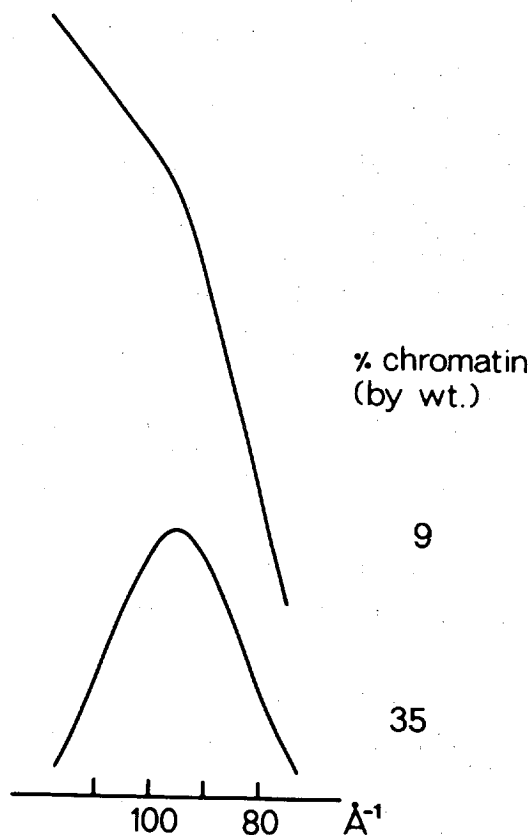


Figure 3. Concentration dependence of 100 Å band in X-ray pattern of chromatin. Chromatin was prepared and X-ray photographs taken as described (Kornberg and Thomas, 1974). The photographs were densitometered and the optical densities radially integrated to obtain the two traces shown.

Electron micrographs of chromatin are also compatible with a jointed structure. After my first talk on the model, at a colloquium in the laboratory last Autumn, Drs A. Klug and J.T. Finch drew my attention to the micrographs of Slayter et al. (1972). These resemble the model, showing thick regions alternating with thin strands, although the distance between thick regions is about 1500 Å. Recently, micrographs have been published (Olins and Olins, 1974) that more closely resemble the model, showing a centre-to-centre distance between thick regions of about 200 Å. There is, of course, still a discrepancy between the micrographs and the model (the distance between thick regions being 100 Å in the model). I believe it is the micrographs which are in error, the distance between thick regions having been extended by partial dissociation of histones and DNA [Olins and Olins mount specimens for microscopy at low ionic strength in the cold, conditions in which the dissociation of F2A2 and F2B is known to occur (Ilyin et al., 1971)].

THE PHASE PROBLEM

The possibility of special sites within the repeating unit where the DNA base sequence is exposed for recognition was mentioned above. Presumably only portions of the DNA base sequence, for example 'receptor' sequences (Davidson and Britten, 1973), need to be exposed in this way. It may be asked how the special sites would come to contain the receptor sequences. There are two main alternatives: (1) there is a specific phase relation between the units and base sequences in the DNA [the phase would be set by non-histone proteins, for example by proteins at the origins of DNA replication; the binding of histones immediately after DNA synthesis (Weintraub, 1974) may be a way of preserving the phase]; (2) the oligomers of histone slide back and forth along the DNA.

ACKNOWLEDGMENT

I thank Drs A. Klug and F. Crick for advice on the X-ray work presented here.

REFERENCES

- Bram, S. and H. Ris. 1971. On the structure of nucleohistone. *J. Mol. Biol.* 55: 325.
- Clark, R.J. and G. Felsenfeld. 1971. Structure of chromatin. *Nature New Biol.* 229: 101.
- Crick, F.H.C. 1971. General model for the chromosomes of higher organisms. *Nature* 234: 25.
- Davidson, E.H. and R.J. Britten. 1973. Organization, transcription, and regulation in the animal genome. *Quart. Rev. Biol.* 48: 565.

- Davies, G.E. and G.R. Stark. 1970. Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. *Proc. Nat. Acad. Sci. USA* 66: 651.
- Ilyin, Y.V., A. Ya. Varshavsky, U.N. Mickelsaar and G.P. Georgiev. 1971. Studies on deoxyribonucleoprotein structure. Redistribution of proteins in mixtures of deoxyribonucleoproteins, DNA, and RNA. *Eur. J. Biochem.* 22: 235.
- Hewish, D.R. and L.A. Burgoyne. 1973. Chromatin substructure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Comm.* 52: 504.
- Kornberg, R.D. 1974. Chromatin structure: A repeating unit of histones and DNA. *Science* 184: 868.
- Kornberg, R.D. and J.O. Thomas. 1974. Chromatin structure: Oligomers of the histones. *Science* 184: 865.
- Luzzati, V. and A. Nicolaieff. 1963. The structure of nucleohistones and nucleoprotamines. *J. Mol. Biol.* 7: 142.
- Noll, M. 1974. Subunit structure of chromatin. *Nature*. In press.
- Olins, A.L. and D.E. Olins. 1974. Spheroid chromatin units (v bodies). *Science* 183: 330.
- Pardon, J.F. and M.H.F. Wilkins. 1972. A super-coil model for nucleohistone. *J. Mol. Biol.* 68: 115.
- Richards, B.M. and J.F. Pardon. 1970. The molecular structure of nucleohistone (DNH). *Exp. Cell Res.* 62: 184.
- Sahasrabudde, C.G. and K.E. van Holde. 1974. The effect of trypsin on nuclease-resistant chromatin fragments. *J. Biol. Chem.* 249: 152.
- Slayter, H.S., T.Y. Shih, A.J. Adler and G.D. Fasman. 1972. Electron microscopy and circular dichroism studies on chromatin. *Biochemistry* 11: 3044.
- van der Westhuyzen, D.R. and C. von Holt. 1971. A new procedure for the isolation and fractionation of histones. *Fed. Europ. Biochem. Soc. Letters* 14: 333.
- Weintraub, H. 1974. The assembly of newly replicated DNA into chromatin. *Cold Spring Harbor Symp. Quant. Biol.* 38: 247.
- Wilkins, M.H.F. 1956. Physical studies of the molecular structure of deoxyribose nucleic acid and nucleoprotein. *Cold Spring Harbor Symp. Quant. Biol.* 21: 75.
- Wilkins, M.H.F., G. Zubay and H.R. Wilson. 1959. X-ray diffraction studies of the molecular structure of nucleohistone and chromosomes. *J. Mol. Biol.* 1: 179.

Origins of chromatin acidic proteins of avian erythroid cells

R. W. H. Harlow and J. R. E. Wells

Nuclear bodies free of membrane as judged by electron microscopy are used as a source of chromatin acidic proteins from avian erythroid cells. For the dividing erythroblast the pattern is relatively complex, whereas in both the non-dividing reticulocyte and transcriptionally inactive erythrocyte the pattern is much simpler, being dominated by three major bands. The acidic protein : DNA ratio decreases with maturation of the cells. Although dispersed chromatin preparations from reticulocytes (purified through 1.7 M sucrose) contain membrane, the acidic protein : DNA ratio is 2.3-fold lower than for reticulocyte nuclear body preparations. However the patterns are similar and it is likely that loss of protein occurs in the former case. Although membrane has been removed, contamination of nuclear body acidic protein preparations with histone and globin seems likely. With these reservations, the results are consistent with the notion that both the amount and complexity of chromatin acidic proteins decrease with the decreasing capacity of cells for transcription.

Current interest in the acidic proteins found in isolated eukaryote chromatin stems from the search for factors within the chromatin that may control gene-specific transcription. Histones are unlikely to be adequate for this task and 'chromosomal RNA' as defined by published preparative procedures (Dahmus and McConnell, 1969; Mayfield and Bonner, 1971), is a product of RNA degradation (Artman and Roth, 1971; Tolstoshev and Wells, 1974).

Cytochemical staining suggests that there are acidic proteins associated with the genetic material *in vivo* (Swift, 1964; Zirkin, 1973), but it is by no means certain that these are adequately represented by the acidic protein populations found in isolated chromatin. The term 'chromatin' is in fact a very loose one, since isolation methods vary enormously and non-specific loss or addition of components can occur. Many chromatin preparations contain membrane material (Jackson et al., 1968; Tata et al., 1972), cytoplasmic proteins (Johns and Forrester, 1969) or RNP-particle proteins (Li and Wang, 1971; Bhorjee and Pederson, 1973), factors that represent results rather than causes of changes in transcription patterns. Both the amount (Dingman and Sporn, 1964; Marushige and Ozaki, 1967; Levy et al., 1972) and pattern (Hill et al., 1971; Seale and

Aronson, 1973; Vidali et al., 1973; LeSturgeon and Rusch, 1973) of chromatin acidic proteins change with transcriptional changes in the cell of origin, or in sub-fractions of the chromatin (Frenster, 1965; Dolbeare and Koenig, 1970). In view of the problems of non-specific contamination of chromatin described above, the significance of these correlative studies is uncertain.

In the avian erythroid series purified populations of cells with defined transcriptional activity may be obtained at different stages of erythroid maturation; from early dividing erythroblasts (highly active in DNA, RNA and protein synthesis) through reticulocytes (non-dividing but active in RNA and globin synthesis) to mature erythrocytes (inactive in macromolecular synthesis). During maturation the decline in transcriptional activity is accompanied by progressive condensation of the chromatin. In this paper we discuss (i) the contribution of membrane to the acidic proteins of isolated chromatin from these cells; (ii) the preparation of membrane-free nuclear bodies from all major cell types and (iii) the analysis of acidic proteins from these preparations in relation to the known transcriptional changes during erythroid maturation.

METHODS

Erythroid cells were obtained and characterised on BSA density gradients as previously described (Harlow et al., 1972).

Chromatin. For erythroid cells, chromatin preparations CC and X were made as previously described (Harlow et al., 1972). Nuclei K were prepared essentially according to Zentgraf et al. (1969), except that homogenisation was performed with a Sorvall Omnimixer at 48,000 rpm, followed by centrifugation through 2.35 M sucrose. SNK nuclear bodies were prepared from saponin-lysed, washed nuclei (Harlow et al., 1972) by similar homogenisation. The outer nuclear membrane was first removed with 0.5% Nonidet P-40, 3 mM CaCl₂; CaCl₂ was included in all subsequent solutions during the preparation.

SNK chromatin was fractionated on Bio Gel A50 columns (70 cm x 1.6 cm) essentially according to the method of Janowski et al. (1972).

Acidic proteins were isolated from chromatin by chromatography on hydroxylapatite (MacGillivray et al., 1971), being eluted with a 0.15 M sodium phosphate step. Histones were similarly prepared as the unbound fraction at 0.001 M phosphate (MacGillivray et al., 1971).

SDS-polyacrylamide gel electrophoresis was by the methods of Weber and Osborn (1969) and Laemmli (1970). Gels were stained and destained by the method of Fairbanks et al. (1971).

Estimations. Protein was estimated according to Lowry et al.

(1951) using BSA as standard, DNA by the diphenylamine reaction (Burton, 1956) and phospholipid (Gehrlach and Deuticke, 1963) and cholesterol (Clark et al., 1968) were measured in chloroform-methanol extracts (Folch et al., 1957).

RESULTS AND DISCUSSION

Transcription in avian erythroid cells As avian erythroid cells mature, progressive condensation of the genetic material (Appels, 1971) is accompanied by a pronounced decrease in levels of transcription measured as incorporation of ^3H -uridine (Figure 1a). Erythroblasts are highly active, reticulocytes exhibit about a 4-fold lower initial incorporation rate, while erythrocytes are essentially inactive.

Isolation of EDTA-particles from the polysomes of reticulocytes pre-labelled with ^3H -uridine shows that there is essentially no ribosomal RNA synthesis whereas a 10S RNA species is labelled (Figure 1b). This RNA has been translated in cell free systems into α and β chicken globin chains (unpublished results). Thus chromatin from reticulocytes and erythrocytes represent templates competent for and unavailable for globin mRNA transcription respectively, and by comparing the chromatin acidic proteins in the two cell types, changes in their composition may be related to the transcriptional difference. Further information may be gained by considering similar comparisons with the earlier cells. It must be emphasised that such studies cannot alone establish a causal connection between chromatin components and the gene-specific control of transcription; they are, however, an essential first step in testing such hypothetical relationships.

No significant changes in either the pattern or amount of histones occur during maturation in these cells (Appels et al., 1972), in particular the mere presence of the tissue-specific histone f2c does not account for the decline in transcription of globin sequences since it is present in comparable amounts in both early and mature cells. Further, the RNA component of chromatin from these cells appears to be a degradation product of larger RNA species, rather than a *bona fide* chromosomal entity (Tolstoshev and Wells, 1974; Harlow et al., 1972). The chromatin acidic proteins do show quantitative and qualitative changes during maturation (see below).

Membrane in avian erythroid chromatin In a previous study (Harlow et al., 1972) we have shown that avian erythroid chromatin from saponin-lysed cells and purified through a 1.7 M sucrose density-barrier (Marushige and Bonner, 1966) still contains significant amounts of membrane.

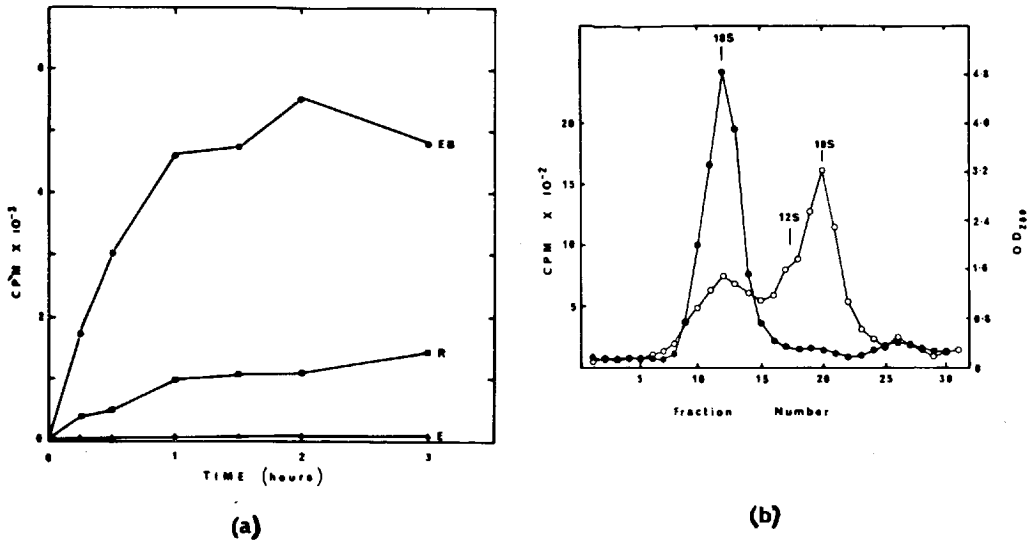


Figure 1. RNA synthesis in avian erythroid cells. The procedures used were basically those of Williams (1971). (a) Cells were incubated with ^3H -uridine at $1 \mu\text{Ci}$ per 2×10^8 cells. Samples were removed at various times, lysed and the TCA insoluble material collected on filters and counted. Symbols: ●—●, erythroblasts (EB); ■—■, reticulocytes (R); ▲—▲, erythrocytes (E). (b) Reticulocytes were pre-labelled with ^3H -uridine and the 20S-30S region of EDTA-treated polysomes isolated (Pemberton et al., 1972). RNA derived from this region was then centrifuged on an SDS-sucrose gradient (SW-41 rotor, 40,000 rpm, 14 hr) and fractions collected. ●—●, $\text{OD}_{260} \text{ m}\mu$; ○—○, cpm.

Analytical data suggested that this contaminant membrane could account for much of the observed chromatin acidic protein. If this were so, one might expect similar proteins in crude chromatin (CC) and the two fractions derived from it after centrifugation on 1.7 M sucrose, namely membrane (M) and 'pure' chromatin (X). The results in Figure 2a show that many such similarities do exist. Furthermore, this situation is not restricted to avian red cells since fractions corresponding to CC, M and X from rat liver chromatin exhibit striking similarities in their acidic protein profiles (Figure 2b). However, when preparations of acidic proteins are separated on discontinuous SDS-gels (Laemmli, 1970) it is clear that some proteins such as j and z are preferentially associated with 'pure' chromatin (Figure 2c) while others such as n and o and minor bands in regions a-g and o-t are probably of membrane origin. To avoid ambiguity it seemed essential to prepare avian chromatin free of

contaminating membrane before the analysis of chromatin acidic proteins.

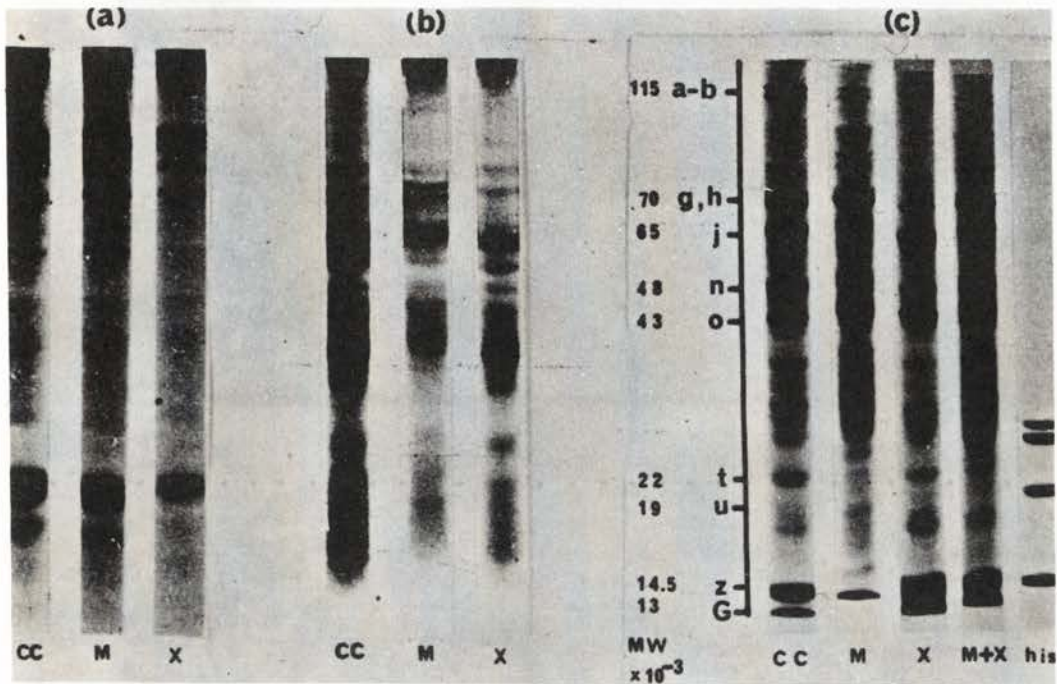


Figure 2. Acidic proteins of chromatins purified on 1.7 M sucrose. Crude chromatin (CC) from chicken reticulocytes (a,c) or rat liver (b) was fractionated on 1.7 M sucrose barriers (Harlow et al., 1972; Marushige and Bonner, 1966) to give membrane (M) and 'pure' chromatin (X) fractions. Acidic proteins were extracted and compared: (a) chicken reticulocyte chromatin fractions and (b) rat liver chromatin fractions, both on continuous buffer, SDS-7.5% acrylamide gels (Weber and Osborn, 1969); (c) chicken reticulocyte chromatin fractions on discontinuous buffer, SDS-acrylamide gels (Laemmli, 1970).

Acidic proteins from membrane-free erythroid chromatin Lysis of avian red cells with either saponin, Nonidet P-40 or Triton X-100 followed by extensive washing in EDTA/NaCl solutions (Harlow et al., 1972) results in nuclear ghost preparations as in Figure 3b. Homogenisation of cells with high speed rotating knives (Kuehl, 1964; Zentgraf et al., 1969) resulted in preparations as seen in Figure 3c. These nuclei (K) maintain similar dimensions to those seen in whole cells, but yields were low (20-30% of DNA recovered as nuclei after sedimentation through sucrose) and there was obvious contamination with haemoglobin (protein/DNA ratio = 4.0). The rotating knives method was then applied to 'saponin' ghosts after removal of the outer nuclear membrane with Nonidet P-40 and 'saponin - Nonidet - knives' or SNK nuclei were obtained in 95% yield and free of membrane (Figure 3d).

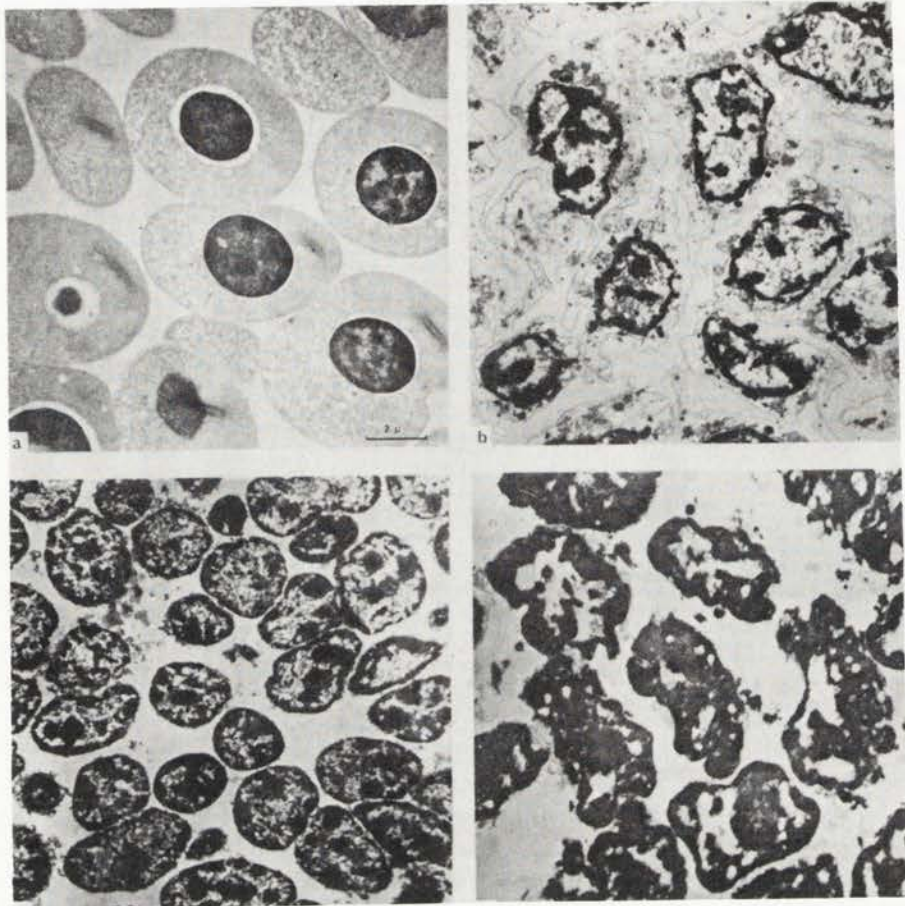


Figure 3. Chicken reticulocyte nuclear preparations. Preparations were fixed, stained, sectioned and examined in a Siemens Elmiskop I electron microscope. (a) whole cells; (b) saponin-lysed cells extensively washed in isotonic sucrose, EDTA, NaCl (Harlow et al., 1972); (c) nuclei K; (d) nuclear bodies SNK.

Various parameters of reticulocyte SNK nuclei are compared with those of CC and X in Table 1. The very low levels of cholesterol and phospholipid in SNK support the observation that membrane is virtually absent in the preparation. The histone:DNA ratio is marginally higher for CC, but is very similar for X and SNK. On the other hand, total protein:DNA is significantly higher for SNK than X, and this is reflected in a 2.3-fold higher ratio of acidic protein:DNA. This is not explained by an increased number of protein

TABLE 1
Composition of erythroid chromatin preparations

Cell	Preparation	$\frac{\text{Protein}}{\text{DNA}}$	$\frac{\text{Histone}}{\text{DNA}}$	$\frac{\text{Acidic protein}}{\text{DNA}}$	$\frac{\text{Cholesterol}}{\text{DNA}}$	$\frac{\text{Phospholipid}}{\text{DNA}}$
Reticulocyte	CC	2.71±0.04	1.59±0.06	0.78±0.02	0.205±0.011	0.614±0.03
	X	1.55±0.02	1.36±0.04	0.30±0.08	0.080±0.006	0.102±0.02
Erythroblast	SNK	1.94±0.21	1.30±0.23	0.69±0.13	0.007±0.010	0.022±0.02
	SNK	2.43±0.15	1.17±0.05	1.25±0.11	-	-
Erythrocyte	SNK	1.62±0.20	1.28±0.17	0.36±0.03	-	-

Chromatin preparations were obtained from the major cell types as described in Methods, and after fractionation on hydroxylapatite (MacGillivray et al., 1971) and extraction of lipids (Folch et al., 1957), estimations were made on at least three independently prepared samples. Ratios are on a weight basis, except for phospholipid: DNA (which is expressed mole/mole of P) and are ± S.E.

species; in fact, the patterns of acidic proteins are comparable. Gels from these preparations, on which equal amounts of protein were loaded, are shown in Figure 4a. Preparation X contains major bands n and o not prominent in SNK; these are probably of membrane origin (see Figure 2c). It is concluded that the lower acidic protein content of X preparations relative to SNK preparations is due to the several homogenisation and centrifugation steps involved in the former preparative procedure (Harlow et al., 1972).

Most of the staining material on gels of SNK acidic protein from reticulocytes is accounted for by bands z (14,500 daltons), α (13,000 daltons), j (65,000 daltons) and g (70,000 daltons). Bands z and α require further comment since they migrate with the same mobility as a histone component and globin (respectively) in this gel system. Major acidic protein components with mobilities identical to histone species in SDS gels are commonly observed (Elgin and Bonner, 1970; Hill et al., 1971). Because acidic protein preparations did not reveal bands on low pH (histone) gels it has been assumed that histones were not present. We also fail to observe such bands even on overloaded gels, but this result is readily explained if aggregation of a histone component with the acidic proteins prevents penetration into the gel. Studies with purified ^{14}C -histone show that contamination of the acidic protein fraction can occur and that this contaminant(s) runs with the same mobility as band z on SDS gels. In addition, preliminary amino acid analysis of band z suggests that it is largely histone.

Because band α has precisely the same mobility as avian globin chains, it is likely that a major proportion of this material is globin. These findings emphasise the problems of assigning chromosomal origins to species isolated as chromosomal acidic proteins.

Comparison of SNK preparations from the three major erythroid cell types shows a significant decrease in acidic protein content as maturation occurs; the ratio acidic protein:DNA falls from 1.25 for the erythroblast, through 0.69 for the reticulocytes, to a final 0.36 in the inactive erythrocyte (Table 1). This difference appears to be due to a progressive decrease in the complexity of the population of acidic proteins (Figure 4b) from a minimum of 27-30 species of widely varying molecular weights in the erythroblast, a less complex pattern in the reticulocyte to the simple population found in the erythrocyte preparation, dominated by the three major bands j, z and α . The significance of two high molecular weight bands in region a (Figure 4b) in the erythrocyte preparation is not known. These findings are consistent with the hypothesis that this class of chromatin proteins are gene-specific regulatory molecules, but equally, the observed changes might be results rather than causes of altered

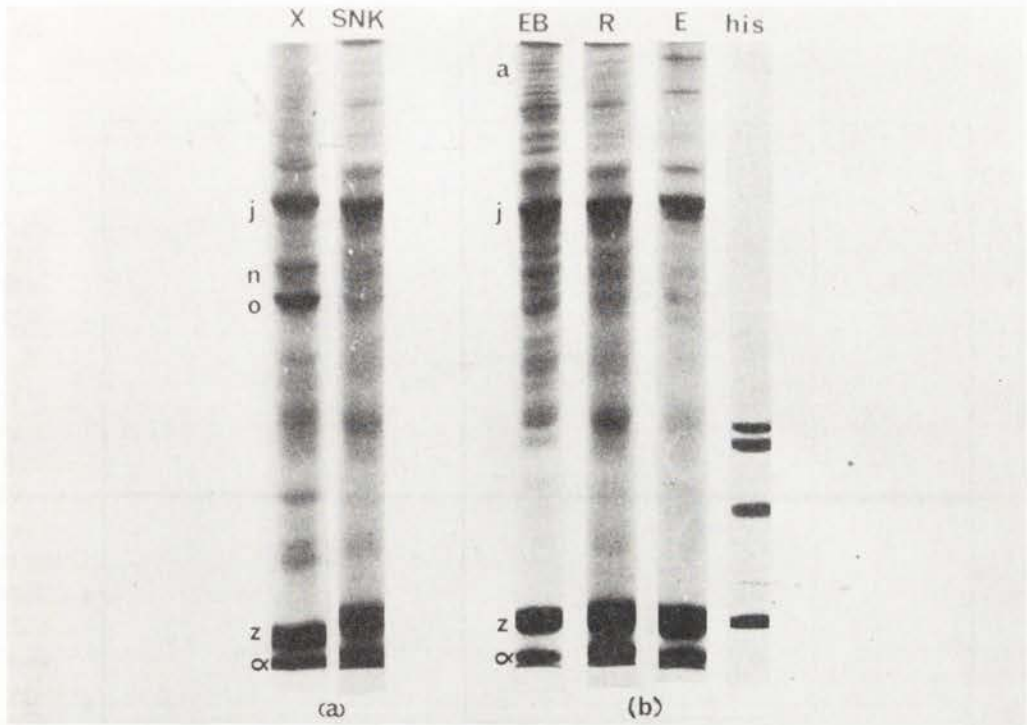


Figure 4. Acidic proteins of SNK nuclear bodies. Acidic proteins were isolated from SNK nuclear bodies of the three major cell types and their analyses on discontinuous SDS-acrylamide gels were compared. (a) comparison of reticulocyte acidic proteins X and SNK; (b) comparison for the different cell types; EB - erythroblast, R - reticulocyte, E - erythrocyte.

gene expression. Molecules involved in transcriptional regulation may be present at levels of only a few molecules per cell (Gilbert and Müller-Hill, 1966), and gel electrophoresis at the present sensitivity may be inadequate to detect them. The large relative amount of species such as band j, as well as its ubiquity throughout maturation (and in other chromatin preparations - Figure 2c) suggests that such a protein may have a structural role analogous to the histones. Despite reservations about bands z and α , preparations of acidic proteins from SNK nuclei probably represent species largely of chromosomal origin, since they are obtained after extensive washing of preparations maintaining nuclear form, but devoid of membrane (Figure 3d; Table 1). Previously published studies of the acidic proteins of isolated avian erythroid chromatin show much more complex patterns for the reticulocyte and erythrocyte (Shelton and Neelin, 1971; Vidali et al., 1973), which are almost certainly due to membrane and possibly cytoplasmic contamination, since in those studies chromatin was prepared merely by saponin-lysis of cells and washing in isotonic sodium chloride (compare Figure 3b). The patterns

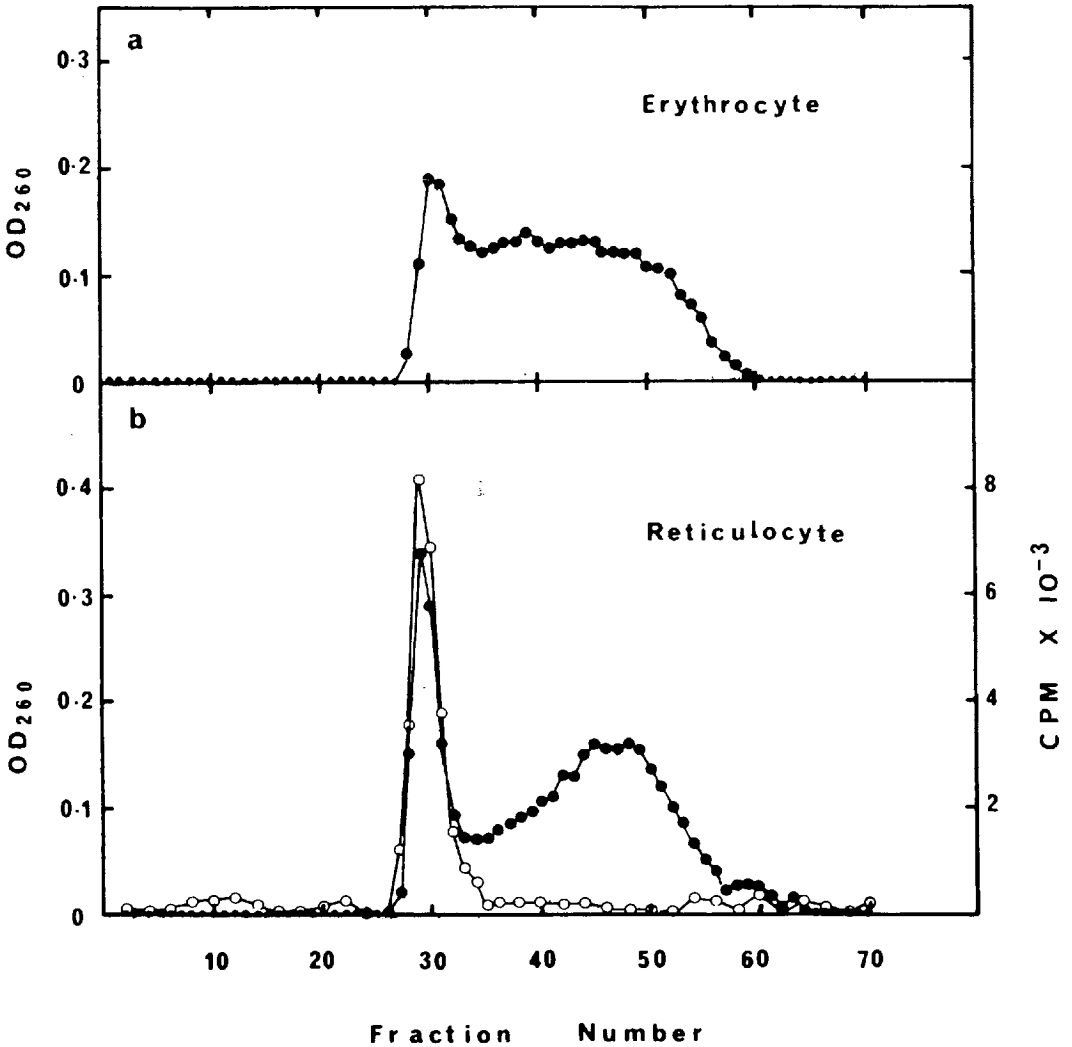


Figure 5. Fractionation of erythroid chromatin on agarose columns. SNK nuclear bodies, prepared as described in Methods, were sheared and fractionated on BioGel-A50 M columns according to Janowski et al. (1972). Reticulocytes were pre-labelled with ³H-uridine as described in Methods. Symbols: ●—●, OD₂₆₀ mμ; ○—○, cpm.

recently reported by Sanders (1974) are much simpler, and approach those obtained for SNK nuclei, though different conditions of electrophoresis were used.

Fractionation of erythroid chromatin While the comparison of chromatin components from reticulocytes and erythrocytes can suggest a possible role in the control of globin mRNA transcription, an assay for *in vitro* transcription of the globin gene is required so that the effects of chromatin components on this process may be tested directly.

Efforts to simplify the chromatin template used in such an *in vitro* assay have led us to attempt fractionation of chromatin. The profile obtained for reticulocyte SNK chromatin (Figure 5) resembles that obtained by Janowski et al. (1972) for mouse liver chromatin. By a number of indirect criteria the leading peak has been assigned as active chromatin (Janowski et al., 1972), and this is consistent with two of our own observations. Firstly, the peak is considerably smaller in the erythrocyte profile, reflecting the lack of transcription and more condensed appearance of the chromatin in this cell. Secondly, as also found by Janowski et al. (1972), when reticulocytes are pre-labelled *in vitro* with ³H-uridine, the labelled RNA appears exclusively in the leading peak. Final confirmation that the leading peak contains the active fraction requires a demonstration of enrichment for globin genes. Our current efforts are directed towards the use of labelled cDNA to verify this and to set up an *in vivo* system to study globin gene transcription.

ACKNOWLEDGMENTS

We thank Dr A.F. Williams and Mr P. Krieg for permission to present their unpublished results, Ms Pamela Y. Dyer for electron microscopy and Ms Jan Dinan for technical assistance. This study was supported by the Australian Research Grants Commission, D65/15758.

REFERENCES

- Appels, R. 1971. The biochemistry of cellular control mechanisms: the characterisation of chromatin from avian erythroid cells. Ph.D. thesis, University of Adelaide.
- Appels, R., J.R.E. Wells and A.F. Williams. 1972. Characterisation of DNA-bound histone in the cells of the avian erythropoietic series. *J. Cell. Sci.* 10: 47.
- Artman, M. and J.S. Roth. 1971. Chromosomal RNA: an artefact of preparation? *J. Mol. Biol.* 60: 291.
- Bhorjee, J.S. and T. Pederson. 1973. Chromatin; its isolation from cultured mammalian cells with particular reference to contamination by nuclear RNP particles. *Biochemistry* 12: 2766.

- Burton, K. 1956. A study of the conditions of the diphenylamine reaction for the colorimetric estimation of deoxyribonucleic acid. *Biochem. J.* 62: 315.
- Clark, B.R., R.T. Rubin and R.J. Arthur. 1968. A new micro-method for determination of cholesterol in serum. *Anal. Biochem.* 24: 27.
- Dahmus, M.E. and D.J. McConnell. 1969. Chromosomal ribonucleic acid of rat ascites cells. *Biochemistry* 8: 1524.
- Dingman, C.W. and M.B. Sporn. 1964. Studies on chromatin. I. Isolation and characterisation of nuclear complexes of DNA, RNA and protein from embryonic and adult tissues of the chicken. *J. Biol. Chem.* 239: 3483.
- Dolbeare, F. and H. Koenig. 1970. Fractionation of rat liver chromatin: effects of cations, hepatectomy and actinomycin D. *Proc. Soc. Exp. Biol. Med.* 135: 636.
- Elgin, S.C.R. and J. Bonner. 1970. Limited heterogeneity of the major nonhistone chromosomal proteins. *Biochemistry* 9: 4440.
- Fairbanks, G., T.L. Steck and G.F.H. Wallach. 1971. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* 10: 2606.
- Folch, J., M. Lees and G.H. Sloane-Stanley. 1957. A simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem.* 226: 497.
- Frenster, J.H. 1965. Nuclear polyanions as de-repressors of synthesis of RNA. *Nature* 206: 680.
- Gehrlach, E. and B. Deuticke. 1963. Eine einfache Methode zur Mikrobestimmung von Phosphat in der Papierchromatographie. *Biochem. Z.* 337: 477.
- Gilbert, W. and B. Müller-Hill. 1966. The *lac* operator is DNA. *Proc. Nat. Acad. Sci. USA* 58: 2415.
- Harlow, R., P. Tolstoshev and J.R.E. Wells. 1972. Membrane - a major source of chromatin - associated RNA and nonhistone proteins in avian erythroid cells. *Cell Diff.* 1: 341.
- Hill, R.J., D.L. Poccia and P. Doty. 1971. Towards a total macromolecular analysis of sea urchin embryo chromatin. *J. Mol. Biol.* 61: 445.
- Jackson, V., J. Earnhardt and R. Chalkley. 1968. A DNA-lipid protein containing material isolated from calf thymus nuclear chromatin. *Biochem. Biophys. Res. Comm.* 33: 253.
- Janowski, M., D.S. Nasser and B.J. McCarthy. 1972. Fractionation of mammalian chromatin. In 5th Karolinska Symposium on Research Methods in Reproductive Endocrinology. Gene Transcription in Reproductive Tissue. (E. Diezfalusy, ed.) p.112. Bogtrykkeriet Forum, Copenhagen.
- Johns, E.W. and S. Forrester. 1969. Studies on nuclear proteins. The binding of extra acidic proteins to deoxyribonucleoprotein

- during the preparation of nuclear protein. *Eur. J. Biochem.* 8: 547.
- Kuehl, L.R. 1964. Isolation of plant nuclei. *Z. Naturforsch.* 196: 525.
- Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680.
- LeSturgeon, W.M. and H.P. Rusch. 1973. Localisation of nucleolar and chromatin residual acidic protein changes during differentiation in *Physarum polycephalum*. *Arch. Biochem. Biophys.* 155: 144.
- Levy, S., R.T. Simpson and H.A. Sober. 1972. Fractionation of chromatin components. *Biochemistry* 11: 1547.
- Li, H.A. and T.Y. Wang. 1971. The ribonucleoprotein particles from nuclear residual proteins of calf thymus. *Arch. Biochem. Biophys.* 144: 549.
- Lowry, O.H., N.J. Rosebrough, A.L. Farr and R.T. Randall. 1951. Protein measurement with the folin phenol reagent. *J. Biol. Chem.* 193: 256.
- MacGillivray, A.J., D. Carrol and J. Paul. 1971. The heterogeneity of the non-histone chromatin proteins from mouse tissues. *Fed. Europ. Biochem. Soc. Letters.* 13: 204.
- Marushige, K. and J. Bonner. 1966. Template properties of liver chromatin. *J. Mol. Biol.* 15: 160.
- Marushige, K. and H. Ozaki. 1967. Properties of isolated chromatin from sea urchin embryo. *Dev. Biol.* 16: 474.
- Mayfield, J.E. and J. Bonner. 1971. Tissue difference in rat chromosomal RNA. *Proc. Nat. Acad. Sci. USA* 68: 2652.
- Pemberton, R.E., D. Housman, H.F. Lodish and C. Baglioni. 1972. Isolation of duck haemoglobin messenger RNA and its translation by rabbit reticulocyte cell-free system. *Nature New Biol.* 235: 99.
- Sanders, L.A. 1974. Isolation and characterisation of the non-histone proteins of developing avian erythroid cells. *Biochemistry* 13: 527.
- Seale, R.L. and A.T. Aronson. 1973. Chromatin - associated proteins of the developing sea-urchin embryo. I. Kinetics of synthesis and characterisation of non-histone proteins. *J. Mol. Biol.* 75: 633.
- Shelton, K.R. and J.M. Neelin. 1971. Nuclear residual proteins from goose erythroid cells and liver. *Biochemistry* 10: 2342.
- Swift, H. 1964. *In The Nucleohistones* (J. Bonner and P.P.P. T'So, ed.) p.169, Holden-Day Inc., San Francisco.
- Tata, J.R., M.J. Hamilton and R.D. Cole. 1972. Membrane phospholipids associated with nuclei and chromatin: melting profile, template activity and stability of chromatins. *J. Mol. Biol.* 67: 231.

- Tolstoshev, P. and J.R.E. Wells. 1974. Nature and origins of chromatin-associated RNA of avian reticulocytes. *Biochemistry* 13: 103.
- Vidali, G., L.C. Boffa, V.C. Littau, K.M. Allfrey and V.G. Allfrey. 1973. Changes in nuclear acidic protein complement of red blood cells during embryonic development. *J. Biol. Chem.* 248: 4065.
- Weber, K. and M. Osborn. 1969. The reliability of molecular weight determinations by dodecyl-sulphate-polyacrylamide gel electrophoresis. *J. Biol. Chem.* 244: 4406.
- Williams, A.F. 1971. Avian erythropoiesis. Ph.D. thesis, University of Adelaide.
- Zentgraf, H., B. Deumling and W.W. Franke. 1969. Isolation and characterisation of nuclei from bird erythrocytes. *Expt. Cell. Res.* 56: 333.
- Zirkin, B.R. 1973. A cytochemical study of the non-histone protein content of condensed and extended chromatin. *Expt. Cell. Res.* 78: 394.

Chromosomal proteins and aspects of the transcription process in eukaryotes

Ronald J. Hill

It is possible to isolate from sea urchin embryos a chromatin complex which retains stage-specific template properties for RNA polymerase. Analysis of the chromatin proteins of the embryos has resolved 11 histones and some 30 nonhistones. Changes occurring in these proteins as a function of development are discussed. Germinal vesicle nuclei were isolated from newt oocytes by microdissection. Centrifugal fractionation of these nuclei gives rise to a nucleoplasm supernatant and a chromatin pellet consisting of lampbrush chromosomes, nucleoli and nuclear membranes. Disulphide bond crosslinks contribute to the morphological integrity of isolated lampbrush chromosomes and nucleoli. The nonhistone proteins from germinal vesicle chromatin exhibit an interesting level of simplicity: they consist predominantly of two major polypeptide components. We may be close to assigning a function to one of these nonhistone species.

It is now over two decades since Stedman and Stedman (1950) advanced the hypothesis that histones may interact with the DNA of cells and thereby influence the phenotype. It is perhaps surprising how little definitive progress has been made since that time towards increasing our understanding of the role played by the chromosomal proteins. However, there is a current upsurge of interest in the assignment of function to proteins of the cell nucleus. I wish to present briefly data bearing on the role played by chromosomal proteins in the transcription process, drawing on two different sources. The first, the sea urchin embryo, is of interest because of the changing pattern of transcription associated with embryonic development. The second, the newt oocyte, promises to be of value because, *inter alia*, of the unusually high level of transcription concomitant with oogenesis.

EXPERIMENTAL PROCEDURES

Embryos of the sea urchin, *Strongylocentrotus purpuratus*, were cultured at 15°C and chromatin isolated at pH 8 in the absence of detergents by the procedures given in Hill et al. (1971). *In vitro* transcription was performed using *Escherichia coli* RNA polymerase as described in Chetsanga et al. (1970). Further experimental details of the macromolecular analysis of sea urchin chromatin may be found in Hill et al. (1971).

Females of the great crested newt, *Triturus cristatus carnifex*, were maintained at 18°C as indicated in Hill et al. (1974). This last communication also gives details of the microdissection operations performed on the germinal vesicle, dissolution of cytological structures and gel electrophoresis of oocyte nuclear proteins.

RESULTS AND DISCUSSION

Transcription and proteins of sea urchin embryo chromatin It has long been established that there is a significant shift in the pattern of transcription during the period of sea urchin development between blastula and pluteus (see e.g. Glisen et al., 1966). Marushige and Ozaki (1967) reported that the template activity of isolated sea urchin chromatin increases between the blastula and pluteus stages. We re-investigated the ability of the embryo chromatin to act as a template for *E. coli* RNA polymerase *in vitro* and reproduced the results of Marushige and Ozaki (Chetsanga et al., 1970). This stage-related increase in template activity could result from (1) an enhanced efficiency in making multiple transcripts of the same active DNA sequences, or (2) transcription of additional DNA sequences.

In order to examine these possibilities we isolated RNA from *in vitro* RNA-polymerase-chromatin reaction mixtures and studied its hybridization properties to denatured DNA at RNA driven C_0t values approaching 50. Such experiments probably largely reflect hybridization to DNA reiterated sequences. When filters loaded with 5 µg of denatured DNA were offered increasing amounts of *in vitro* synthesized RNA, hybridization plateaus were reached at 100 µg RNA input (Figure 1). RNA synthesized on naked DNA gave a plateau corresponding to some 20% of the DNA hybridized. RNA transcribed from chromatin yielded considerably lower plateaus suggesting that certain regions of the genome were genetically silent. Blastula-chromatin-templated RNA formed hybrid with 4% of denatured DNA while RNA transcribed from pluteus-chromatin hybridized with 8% of the DNA. These findings are consistent with the notion that more of the genome is available for transcription in pluteus chromatin than in blastula chromatin.

Evidence of this type suggests that there is, incorporated into the structure of the isolated chromatin complex, information encoding a significant shift in the pattern of transcription between blastula and pluteus. Experiments described elsewhere (Chetsanga et al., 1970) further imply that this shift is related to *in vivo* events occurring during embryonic development.

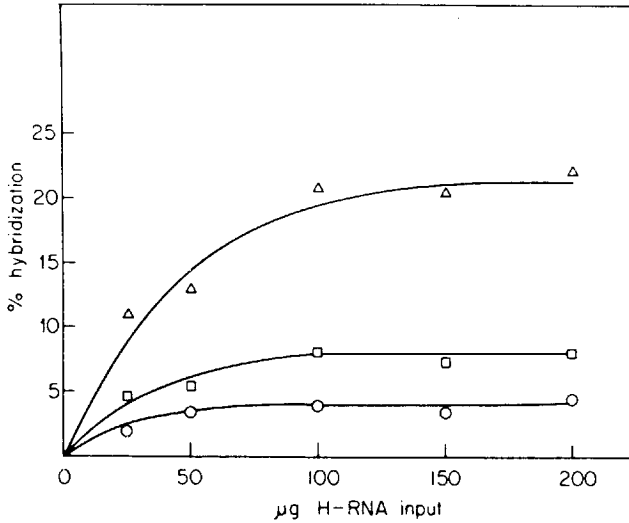


Figure 1. Hybridization of RNA made from sperm DNA (Δ — Δ), pluteus chromatin (\square — \square) and blastula chromatin (\circ — \circ), to sperm DNA. Incubations were performed at 37° in 1 ml of 30% formamide, 2 x SSC. The washing procedure included RNase treatment.

Let us now turn to the chemical nature of this chromatin complex. At the time when we were initiating our studies the methods available in the literature for resolution of chromatin into its macromolecular components were subject to a number of shortcomings. I therefore decided to develop independent techniques.

The interactions between the components of chromatin - DNA, RNA, histones and nonhistones - must surely involve hydrogen, hydrophobic and ionic bonds. Guanidine hydrochloride, thought to break all three classes of noncovalent interactions (Tanford, 1964), might well be expected to be an ideal dissociating agent for chromatin. Figure 2a depicts the elution pattern of sea urchin gastrula chromatin dissociated in 4 M guanidine hydrochloride at pH 8 on gel filtration through a column of Sepharose 4B. A high-molecular-weight peak (F_H) of relatively pure nucleic acid appears close to the void volume and well resolved from a lower-molecular-weight predominantly protein peak (F_L).

In order to simplify the interpretation of electrophoretic analyses and to minimize the possibility of electrostatic interactions I decided to separate next the proteins in F_L into positive and negative charge classes. The rationale was to maximize dissociating conditions and to employ the strongest potential

extrinsic interaction, that between histones and a negatively-charged ion-exchange matrix, to remove one charge class from solution.

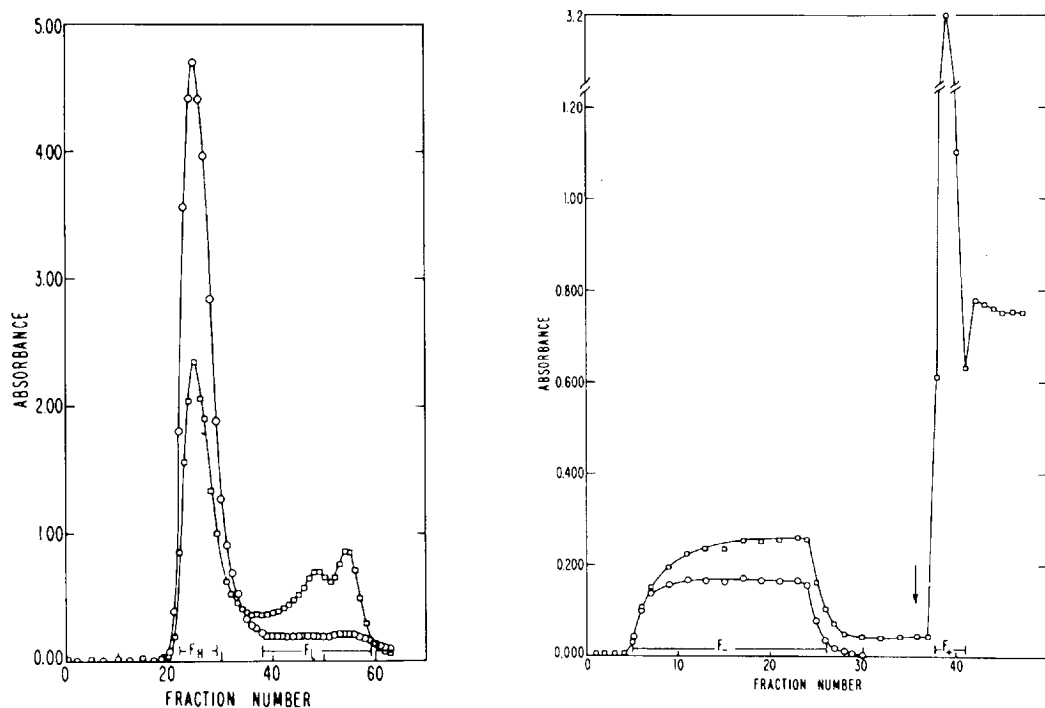


Figure 2 (a) Elution profile of gastrula chromatin in 4 M guanidine hydrochloride on Sepharose 4B. (b) Charge class separation of gastrula F_L on carboxymethyl Sephadex. At the point indicated by the arrow the high ionic strength eluant was introduced.

A_{260} . \circ — \circ : A_{230} . \square — \square

Of the many systems tested, the best, both from the viewpoint of charge class separation and subsequent recovery of histones, was carboxymethyl Sephadex equilibrated with 6 M urea, 0.37 M guanidine hydrochloride, 0.01 M sodium pyrophosphate at pH 7.8. On passing gastrula F_L through such a column approximately half of the A_{230} units are unretarded (Figure 2b). On subsequent elution with 40% (w/v) guanidine hydrochloride a second fraction is eluted as a sharp peak. The material running straight through the column might be expected to bear a net negative charge and is thus termed F_- . The material bound and eluted at high ionic strength might be expected to be positively charged and is termed F_+ . Both fractions F_- and F_+ were reduced and alkylated prior to electrophoresis by the procedure of Crestfield et al. (1963).

Electropherograms of fractions F_+ from blastula, gastrula and pluteus at pH 3 in 2.5 M urea reveal characteristic histone patterns (Figure 3). It can be seen that the five major histone classes are present in sea urchin embryo chromatin and that these are further resolved to reveal fine structure. Stage-specific quantitative changes are apparent. There is a progressive increase in components 1' and 4' while there is a progressive decrease in 2''' with increasing time of development.

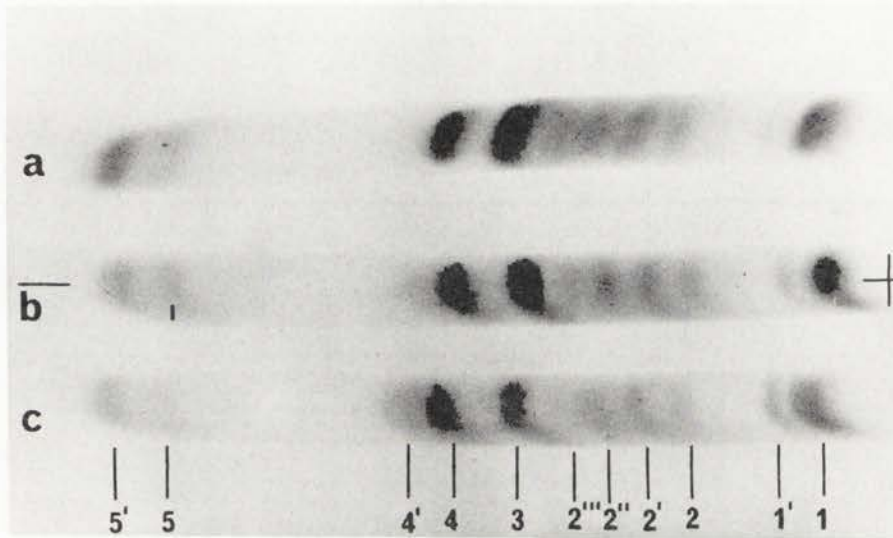


Figure 3. Acrylamide gel electropherograms in 2.5 M urea, pH 3. (a) Blastula F_+ ; (b) gastrula F_+ ; (c) pluteus F_+ . The numbers 1 to 5 refer to the five common histone classes.

Fractions F_- from blastula, gastrula and pluteus in the presence of 0.1% sodium dodecyl sulphate (SDS) at pH 7 gave the electropherograms recorded in Figure 4. A large number of acidic chromatin proteins are resolved. There is an overall similarity in the patterns given by the different developmental stages. However, close examination of the electropherograms reveals definite changes in proportions of individual components. The most striking change is the decrease in the component designated A27 from a strong band at blastula to a barely perceptible one at pluteus. There is also a decrease in A48 and significant increases in components A78 and A103 with increasing time of development.

Those protein components exhibiting the most dramatic changes during development are clearly eligible for a role in the regulation of genetic expression. However, the overall qualitative

similarities in the chromosomal protein populations, in the face of the significant shift in transcription pattern between the early embryonic blastula and the much later pluteus when organogenesis is well established, suggest that many of the proteins may serve more universal roles. Obvious possibilities are structural roles in the chromatin and nuclear membrane, enzymes and processors of RNA such as 'informatin' (Krichevskaya and Georgiev, 1969).

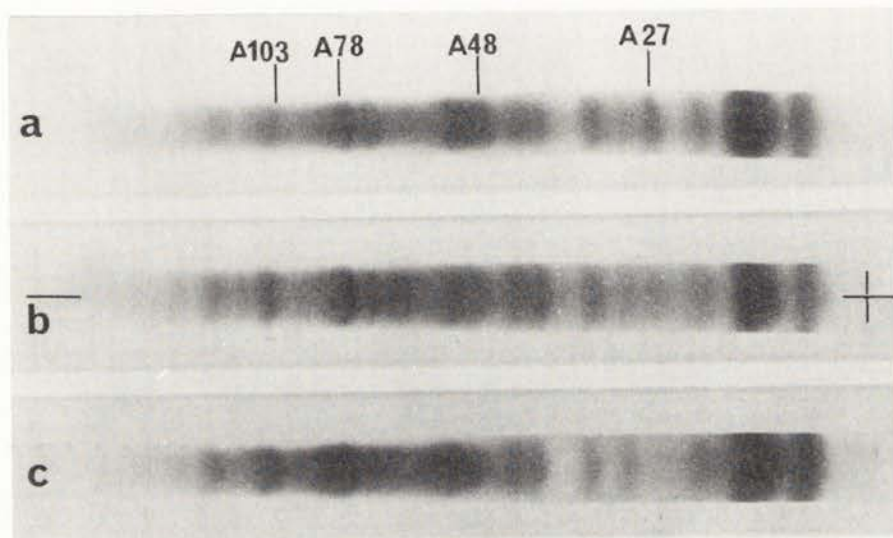


Figure 4. Gel electropherograms in 0.1% SDS, pH 7. (a) Blastula F_- ; (b) gastrula F_- ; (c) pluteus F_- . Bands exhibiting more obvious changes have been designated by the prefix A (for acidic) followed by the apparent molecular weight in kilodaltons.

Proteins of the oocyte nucleus of the newt The newt germinal vesicle offers a potentially interesting focus of inquiry. This giant oocyte nucleus may be isolated by microdissection in an operation monitored directly under the microscope and performed much more rapidly than classical bulk biochemical separations. Furthermore, there is a vast body of cytological information available; within the isolated nuclei peripheral nucleoli and a central mass of lampbrush chromosomes are demonstrable by phase microscopy (Callan and Lloyd, 1960).

Removal of the nuclear envelope releases the lampbrush chromosomes, nucleoplasmic granules and a variable proportion of the nucleoli. A typical portion of such a preparation from *Triturus cristatus carnifex* after centrifugation is illustrated in Figure 5a. In an attempt to achieve molecular dispersion of these structures, aqueous guanidine hydrochloride, buffered at pH 8.0,

was added to a final concentration of 4 M. The optical contrast of the preparation decreased considerably (Figure 5b), but all cytological structures of the original preparation were still apparent.

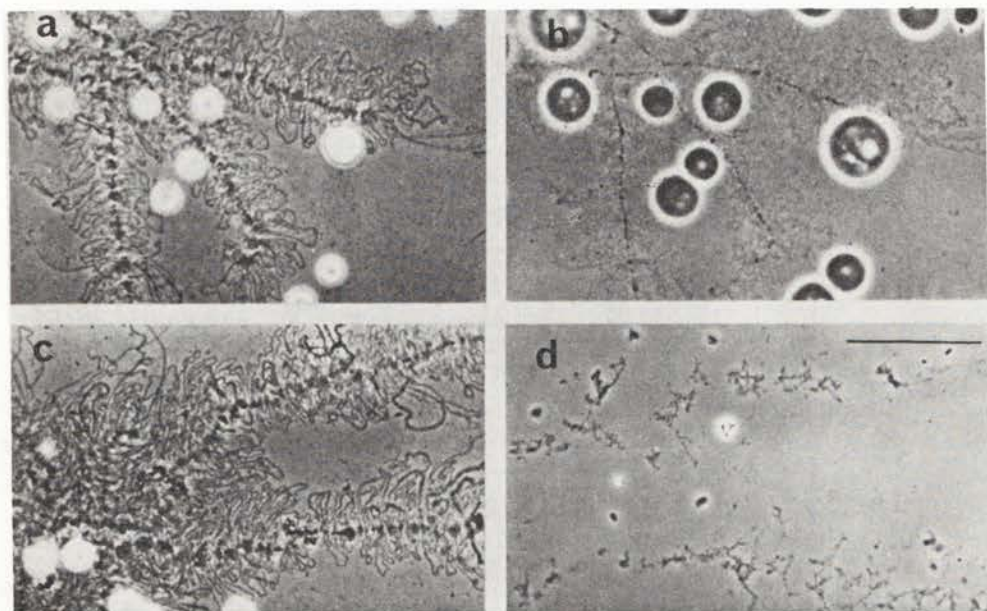


Figure 5. Effects of guanidine hydrochloride and disulphide reducing agents on *Triturus* lampbrush chromosomes and nucleoli. Before (a) and after (b) treatment with 4 M guanidine hydrochloride; before (c) and after (d) treatment with guanidine hydrochloride in 0.1 M mercaptoethanol. The bar represents 50 μM .

A far more dramatic effect is observed when this experiment is repeated with the modification of incorporating 0.1 M mercaptoethanol in the final guanidine solvent (Figures 5c and d). Dissolution has gone to near completion and the original cytological structures have collapsed. Traces of cores are all that remain of the nucleoli. The chromosome axes can be traced as barely recognizable strings, no doubt held together by the DNA backbone. Replacing mercaptoethanol with dithiothreitol or with thioglycollate also leads to near complete dissolution of the structures (not shown).

All three of the reagents utilized (mercaptoethanol, dithiothreitol and thioglycollate) are well established as specific reagents for the reductive cleavage of disulfide bonds. These observations thus provide circumstantial evidence that disulfide bond crosslinks contribute to the morphological stability of isolated nucleoli and lampbrush chromosomes.

In order to obtain proteins for electrophoretic examination, hand-isolated germinal vesicles were massed and fractionated into nucleoplasm and chromatin. Their membranes were allowed to rupture spontaneously on standing in a solution of 0.075 M KCl and 0.025 M NaCl. After gentle mixing, centrifugation at 1000 g for 10 min gave a nucleoplasm supernatant over a pellet of chromatin.

The unusually high protein to nucleic acid ratio of the germinal vesicle allows the following simplified procedure for analysis of nuclear proteins. Nuclear proteins were prepared from total nuclei, nucleoplasm and washed chromatin by treatment with 2 μ g/ml pancreatic ribonuclease in 2 M urea (introduced as a precaution to degrade RNA and minimize any possible RNA-protein interactions during electrophoresis), followed by dissolution of all structures in 4 M guanidine hydrochloride, 0.1 M mercaptoethanol. Finally, the proteins were S-carboxymethylated essentially by the procedure of Crestfield et al. (1963). Gel electropherograms at pH 9 in 7 M urea are depicted in Figure 6. Total nuclear protein is resolved into some twelve bands. Of these, one (1.30) is very much depleted in the nucleoplasmic fraction. A second (2.65) appears to be present in slightly decreased proportions in the nucleoplasm. Both of these components are very much enriched in the chromatin fraction. The mobility of these two species towards the anode at pH 9 eliminates them from the histone class.

The total chromatin pellet obtained by low-speed centrifugation of ruptured whole nuclei consists of nuclear membranes, nucleoli and lampbrush chromosomes (Figure 7a). At early stages of oogenesis the nucleoli are attached to the nuclear membrane. Thus, if prior to the collection of nuclear material at early stages of oogenesis, the nuclear membranes are removed, the nucleoli are removed along with them so that the washed chromatin pellet contains lampbrush chromosome material and little else (Figure 7b). The membranes and attached nucleoli may also be massed and washed thoroughly to give material corresponding to that illustrated in Figure 7c. Finally, isolated membranes with very few attached nucleoli may be obtained from oocytes at later stages of oogenesis (Figure 7d).

The protein pattern given by whole nuclei is depicted in Figure 8a. Total chromatin, displaying the two enriched components (1.30 and 2.65), is represented by Figure 8b. Both of these enriched species are again present in the pelleted membranes with attached nucleoli (Figure 8c). A corresponding amount of membrane material largely free from nucleoli gave no significant electropherogram (Figure 8d). We must therefore conclude that the two

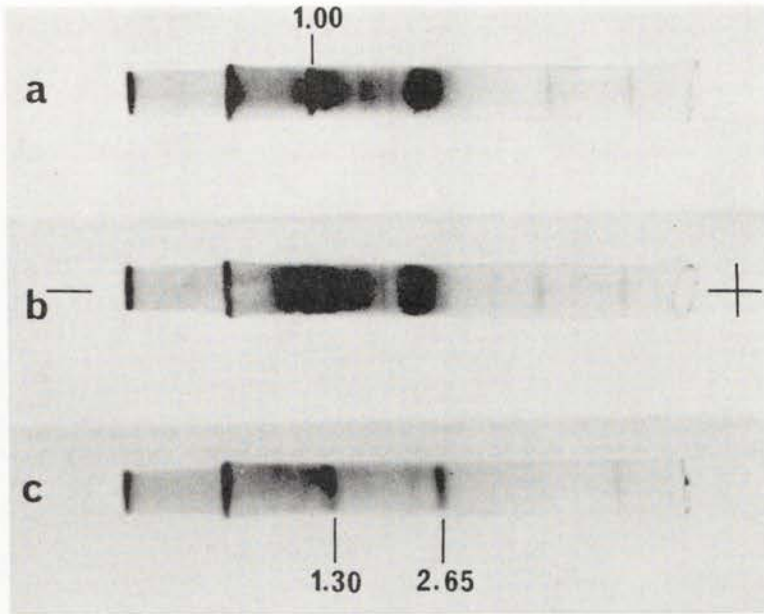


Figure 6. Gel electropherograms given by solubilized, reduced and alkylated germinal vesicle proteins: (a) nucleoplasmic proteins; (b) total nuclear proteins; (c) chromatin proteins. The numbers associated with the bands indicate relative mobilities. Electrophoresis according to Davis (1964) with the inclusion of 7 M urea.

major species are stemming, at least in considerable part, from the nucleoli. Figure 8c records the pattern given by lampbrush chromosomes alone. Only one faint band, in position 1.00, could be discerned on the original gel. Microscopic observations indicate that unprotected lampbrush chromosomes are exceedingly fragile structures with a marked predilection for shedding components on the application of mild shear forces. More work must be done to produce bulk quantities of isolated lampbrush chromosome material.

Both basic and acidic proteins adopt a high net negative charge on exposure to SDS. On electrophoresis in the presence of this detergent the total chromatin protein pattern again differs markedly from the whole nuclear pattern and once more demonstrates the presence of the two distinct major components (Figure 9). The molecular weights of these major components, calculated from SDS gel electrophoresis, are 110,000 and 43,000. The region where histones would be expected to run is indicated by the bracket. This class of proteins thus, at most, makes only a minor contribution to the overall protein population. The two major non-

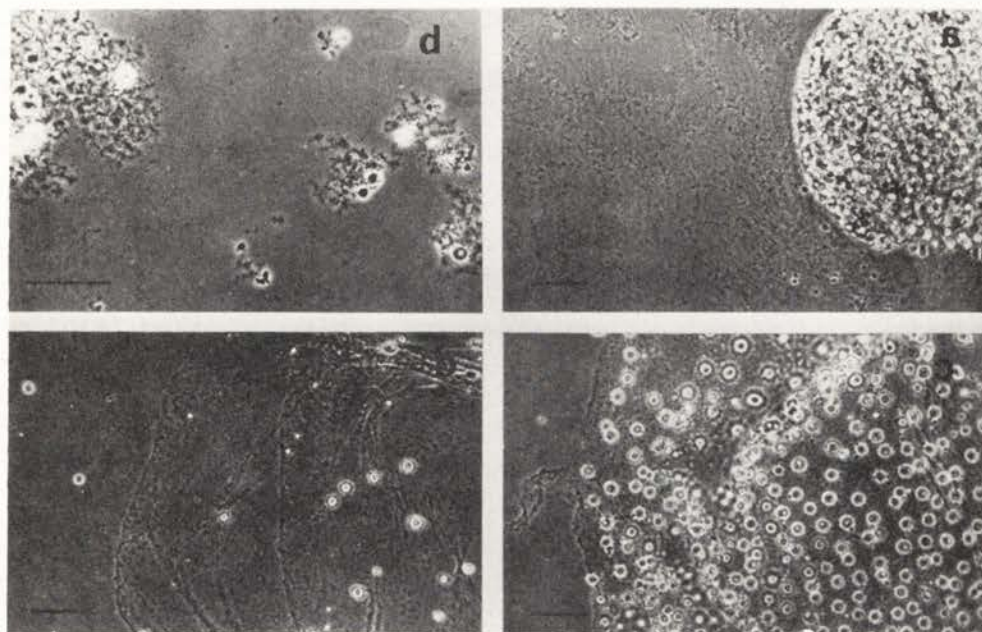


Figure 7. Oocyte chromatin and its isolated subfractions. (a) Total chromatin; (b) bulk lampbrush chromosomes; (c) nuclear membrane with attached nucleoli; (d) membrane from older oocytes with very few attached nucleoli. The bars correspond to 50 μm .

histones therefore predominate over all other proteins.

This level of simplicity in the oocyte chromatin protein population is outstanding. We may have a clue as to the role of at least one of the major nonhistones. An unusual feature of oocyte chromatin (both nucleoli and lampbrush chromosomes) is the high content of RNA relative to DNA (Macgregor, 1972; Edstrom and Gall, 1963). In fact, the enrichment of the RNA to DNA ratio compared with somatic chromatin (see e.g. Hill et al., 1971) is of the order of ten to one hundredfold. Thus much of the protein of germinal vesicle chromatin is likely to be associated with nascent RNA. In other words, nascent chromosomal RNA-associated proteins are likely to be amplified relative to the other non-histones of somatic chromatin. At this point it may be germane to recall the very simple pattern given by fully dissociated 'informatin' protein in the hands of Krichevskaya and Georgiev (1969). This protein, which is believed to associate with nuclear

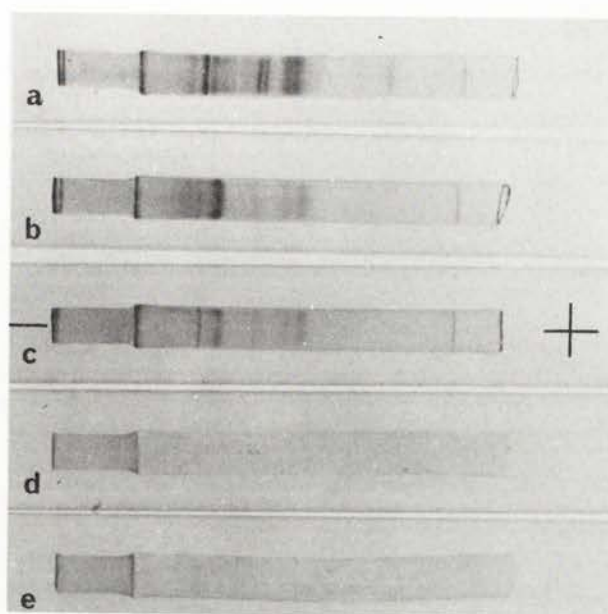


Figure 8. Urea gel electropherograms of S-carboxymethylated proteins prepared from (a) whole nuclei; (b) total chromatin; (c) membranes with nucleoli; (d) membranes alone; and (e) isolated chromosomes.

RNA and be involved in its processing, gives a single band on gel electrophoresis. Its molecular weight has been quoted as 40,000-45,000. It may be relevant that informatin polymerizes via disulfide bond formation.

Somatic chromatin complexes, in which the RNA is a minor component relative to DNA, generally give a complex pattern of nonhistones (see e.g. Shelton and Allfrey, 1970; Hill et al., 1971). However, when RNA synthesis in rat liver cells is stimulated by cortisol there is a concomitant response in a single chromatin acidic protein. This species has a molecular weight of 41,000 (Shelton and Allfrey, 1970). On ecdysone-induced puffing of *Drosophila* polytene chromosomes there occurs an accumulation of a specific nonhistone of molecular weight 42,000 (Helmsing and Berendes, 1971). Furthermore, Bhorjee and Pederson (1973) have observed that when mammalian chromatin is prepared by a novel procedure which halves its content of RNA, the content of a particular protein of approximate molecular weight 40,000 is also halved. There is an interesting possibility that the nonhistone species from newt germinal vesicles, rat liver and *Drosophila*

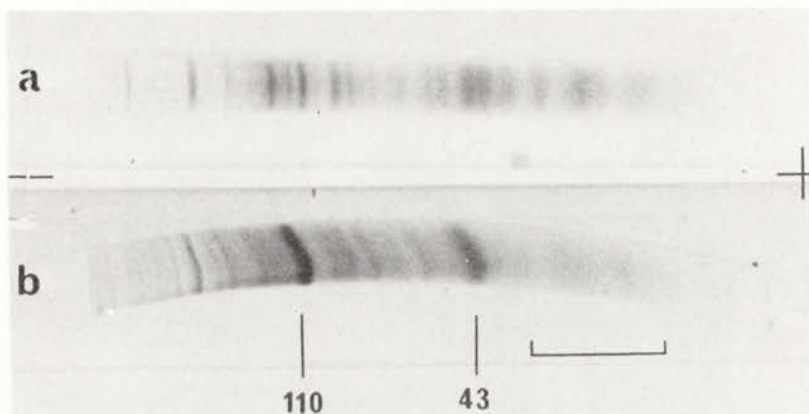


Figure 9. SDS gel electropherograms of S-carboxymethylated proteins from (a) whole germinal vesicles and (b) total chromatin. The bracket delineates the region where the five common histones would be expected to run. Electrophoresis according to Shapiro et al. (1967).

polytene nuclei, all with molecular weights measured to be in the range 41,000-43,000 and all associated with stimulated transcription, may represent information functioning at the chromosomal level.

ACKNOWLEDGEMENTS

The sea urchin work was performed in collaboration with C.J. Chetsanga, D.L. Poccia and Paul Doty at Harvard University. The newt oocyte data were obtained in collaboration with H.G. Callan and K. Maundrel at the University, St. Andrews. The formal references that I have made are at best a very inadequate representation of the efforts of my co-workers.

The figures have been reproduced by permission from the publishers. Figure 1 is reproduced from Chetsanga et al. (1970); Figures 2 to 4 are from Hill et al. (1971); Figures 5 to 9 are from Hill et al. (1974). Figures 5, 6 and 9 also appeared in Hill, Maundrel and Callan (1973) "Molecular Cytogenetics" (B.A. Hamkalo and J. Paparonstantinou, eds.) p.147, Plenum Press, New York.

REFERENCES

- Bhorjee, J.S. and T. Pederson. 1973. Chromatin: Its isolation from cultured mammalian cells. *Biochemistry* 12: 2766.
 Callan, H.G. and L. Lloyd. 1960. Lampbrush chromosomes of crested newts. *Phil. Trans. Roy. Soc. London B* 243: 135.

- Chetsanga, C.J., D.L. Poccia, R.J. Hill and P. Doty. 1970. Stage-specific RNA transcription in developing sea urchins and their chromatins. *Cold Spring Harbor Symp. Quant. Biol.* 35: 629.
- Crestfield, A.M., S. Moore and W.H. Stein. 1963. The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. *J. Biol. Chem.* 238: 622.
- Davis, B.J. 1964. Disc electrophoresis - II. Method and application to human serum proteins. *Annals New York Acad. Sci.* 121: 404.
- Edstrom, J.E. and J.G. Gall. 1963. The base composition of RNA in lampbrush chromosomes, nucleoli, nuclear sap and cytoplasm of *Triturus* oocytes. *J. Cell Biol.* 19: 279.
- Glisen, V.R., M.V. Glisen and P. Doty. 1966. The nature of m-RNA in the early stages of sea urchin development. *Proc. Nat. Acad. Sci. USA* 56: 285.
- Helmsing, P.J. and H.D. Berendes. 1971. Induced accumulation of nonhistone proteins in polytene nuclei of *Drosophila hydei*. *J. Cell Biol.* 50: 893.
- Hill, R.J., K. Maundrel and H.G. Callan. 1974. Nonhistone proteins of the oocyte nucleus of the newt. *J. Cell Sci.* In press.
- Hill, R.J., D.L. Poccia and P. Doty. 1971. Towards a total macromolecular analysis of sea urchin embryo chromatin. *J. Mol. Biol.* 61: 445.
- Krichevskaya, A.A. and G.P. Georgiev. 1969. Further studies on the protein moiety in nuclear DNA-like RNA containing complexes. *Biochim. Biophys. Acta* 164: 619.
- Macgregor, H.C. 1972. The nucleolus and its genes in amphibian oogenesis. *Biol. Rev.* 47: 177.
- Marushige, K. and H. Ozaki. 1967. Properties of isolated chromatin from sea urchin embryo. *Devel. Biol.* 16: 474.
- Shapiro, A.L., E. Vinuela and J.V. Maizel, Jr. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem. Biophys. Res. Comm.* 28: 815.
- Shelton, K.R. and V.G. Allfrey. 1970. Selective synthesis of a nuclear acidic protein in liver cells stimulated by cortisol. *Nature* 228: 132.
- Stedman, E. and E. Stedman. 1950. Cell specificity of histones. *Nature* 166: 780.
- Tanford, C. 1964. Cohesive forces and disruptive reagents. In *Subunit Structure of Proteins*. Brookhaven Symposia in Biology, Vol. 17 p. 154.

Zygotene and pachytene-labeled sequences in the meiotic organization of chromosomes

Yasuo Hotta and Herbert Stern

Single-copy and repeated DNA sequences are synthesized during the prophase of meiosis in *Lilium* microsporocytes. The single-copy sequences are replicated late during zygotene (Z-DNA) but are not ligated to the remainder of the DNA until some time between the completion of pachytene and the termination of meiosis. The Z-DNA regions are estimated to be about 3.5 kilobases (kb) in length and the intervals between them to be about 10^3 kb. Repeated DNA sequences are synthesized at pachytene (P-DNA) and are sites of repair replication. The average length of these sites is estimated as 170 bases. They are classified as moderate repeats since they are reiterated 2×10^3 times per genome and have a complexity of the order of 5×10^4 base pairs. The single-copy sequences are speculatively assigned a function in synapsis and the repeated sequences a function in crossing-over.

Two forms of synapsis occur during meiosis, one observable by light or electron microscopy as a pairing of chromosomes, and the other inferred from the molecular requirements for DNA recombination. The relationship between these forms has been variously interpreted although circumstantial evidence from a variety of organisms supports the view that recombinational pairing follows the prior establishment of chromosome synapsis (Stern and Hotta, 1973). Whether or not this is so, a pertinent question is whether there are elements of chromosome organization that are specifically designed to effect or facilitate these two forms of synapsis. The *Lilium* genome, for example, consists of about 30 m of duplex DNA. During meiosis, 120 m of the DNA are confined to a nucleus of approximately 10 μ m diameter, and yet crossing-over between homologs occurs regularly with a frequency of 2-3 chiasmata in each of the 12 bivalents. Such regularity in crossing-over requires a corresponding regularity in the effective juxtaposition of homologous DNA strands. It is generally presumed that special mechanisms exist to enable homologous strands to find each other and to pair at least long enough for recombination to be completed. Since gene homology is a primary requirement of recombinational pairing, the alignment between homologs must occur with a high degree of precision. It is difficult to conceive of regulating such alignment in the packed *Lilium* nucleus solely by chance collisions between corresponding stretches of DNA.

If chromosomes do have specific structures concerned with alignment, then analyses of meiotic chromosome behavior should reveal such aspects of chromosome organization. Although studies in our laboratory have been addressed primarily to metabolic activities during meiotic prophase, they do, in fact, provide pointers on the nature of chromosome organization, particularly with respect to the mechanics of pairing and crossing-over. These pointers are of limited scope but they are nevertheless helpful in designing models of meiotic chromosomes and in relating such models to chromosome organization in general.

In this paper we describe experiments that help to clarify the relationship between two distinctive but very small fractions of DNA in *Lilium* and the genome as a whole. Specifically, we will show that the satellite properties of the DNA synthesized during zygotene (Z-DNA) are a transient characteristic and that Z-DNA actually represents stretches of single-copy sequences. We will also attempt to show that the repeat sequences of DNA synthesized during pachytene (P-DNA) are distributed among all classes of DNA in the nuclear genome.

METHODS

Procedures used in handling *Lilium* microsporocytes are the same as those described previously (Stern and Hotta, 1970).

DNA was prepared from isolated nuclei by the organic solvent-pronase method described previously by Hotta and Bassel (1965). The purified DNA was dissolved together with ^{14}C -*Escherichia coli* DNA in 0.12 M phosphate buffer, pH 6.8, and sheared by sonication to a mean chain length of 450-600 nucleotides. Samples were denatured by heating to 100°C for 20 min. Reassociation was carried out in sealed ampules at 65°C. C_0t values were obtained with hydroxyapatite columns following the procedure described by Britten and Kohne (1968). For values of C_0t greater than 5×10^3 mole sec/liter, the reassociations were carried out in the presence of 1.5 M NaCl. Data obtained in this way have been normalized to standard phosphate conditions. Rates of reassociation in the presence of NaCl were approximately 8 times those in its absence.

Additional experimental details are included in legends to figures and tables.

RESULTS

The organization of zygotene-labeled sequences Our speculations on various cellular and molecular aspects of chromosome synapsis have been more than adequately discussed (Stern and Hotta, 1973, 1974). Their essentials are here summarized to provide a background for the results to be described:

(1) Zygotene chromosomes undergo pairing as appreciably condensed structures with no more than about 1% of their DNA being exposed for interactions between homologs.

(2) Since the pairing between homologs is genetically precise in normal meiotic situations, matching of partly condensed chromosomes must be achieved by relatively short intercalary stretches of DNA. These are distributed throughout the length of each homolog and remain exposed for interaction after partial condensation of the chromosomes. The regions of interaction are termed 'synaptic sites'.

(3) Z-DNA is an attractive candidate as a synaptic agent since its delayed replication coincides with and appears to be essential to chromosome pairing. Repetitious DNA is an unattractive candidate because it might promote poor alignment between homologous chromosomes, and, if broadly distributed, it would also promote pairing between heterologous ones.

(4) The geometry of the synaptonemal complex and its associated chromosomes is such as to suggest that DNA-DNA matching between synaptic sites is a transient event and that matched segments are immediately modified and stabilized by formation of the synaptonemal complex.

(5) The synaptonemal complex is separate and separable from the process of juxtaposing synaptic sites even though its formation is usually associated with gene-specific synapsis. The specificity is a consequence of temporal co-ordination between DNA-DNA matching and synaptonemal complex formation. The synaptonemal complex can and does form between non-homologous strands (Gillies, 1973; Menzel and Price, 1966).

In early biochemical studies of meiosis in *Lilium*, we observed that an identifiable fraction of nuclear DNA was synthesized during zygotene (Hotta et al., 1966). This late replicating Z-DNA was found to be essential to chromosome synapsis inasmuch as inhibition of its replication arrested cells at zygotene. Monitoring of cells by electron microscopy further indicated that DNA synthesis and synaptonemal complex formation were tightly coupled (Roth and Ito, 1967). The relationship has been observed in all liliaceous plants tested and evidence for its occurrence in rat spermatocytes will be described later.

In addition to its delayed synthesis, Z-DNA is distinguished by its banding in CsCl gradients as a satellite with a buoyant density of 1.712 g/cc. Total nuclear DNA has a buoyant density of 1.702 g/cc (Hotta and Stern, 1971). Although the satellite, which constitutes about 0.3% of the total DNA, is tracked by radioactivity, chemical analysis of its base composition matches that calculated from its

mean buoyant density. The correspondence permits the conclusion that Z-DNA represents a discrete class of GC-rich molecules. Superficially at least, its separate banding and distinctive composition are typical for satellite DNA. A conclusion which might be drawn from these properties is that Z-DNA represents a class of highly repeated sequences which functions in chromosome synapsis. Such a role has indeed been proposed for heterochromatin (Yunis and Yasmineh, 1971).

As mentioned earlier, despite its satellite-like behavior, Z-DNA is not a collection of repeated sequences. C_{0t} analysis of Z-DNA is the basis for this conclusion. DNA labeled during zygotene begins to reassociate only after 60% of the genome as a whole has reassociated (Figure 1). The principal portion of the Z-DNA reassociation curve does not parallel the profile for total DNA even in the high C_{0t} region. Since all the Z-DNA sequences replicated semi-conservatively during the zygotene stage, the reassociation reaction is between labeled fragments and is not driven by unlabeled sequences. Thus, from the standpoint of reassociation kinetics, Z-DNA represents a distinctive fraction of the genome. As a further check on its behavior, the Z-DNA label was separated from bulk nuclear DNA by two different procedures and each preparation analyzed

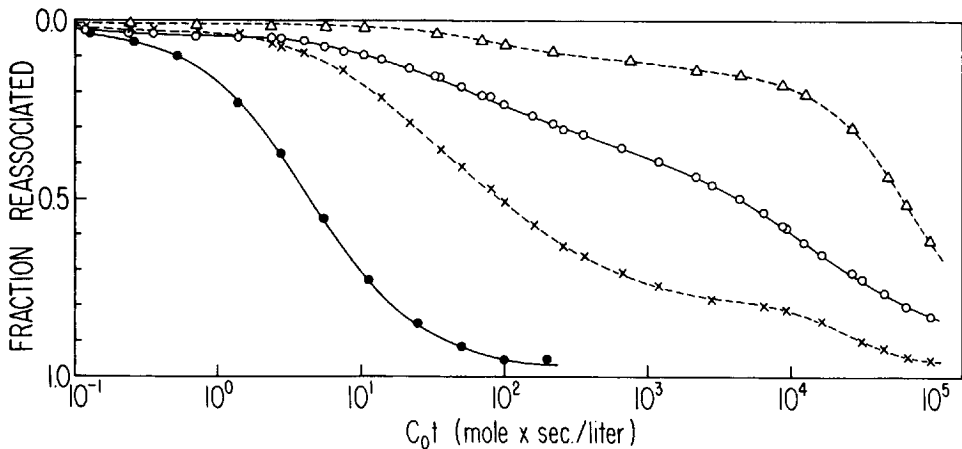


Figure 1. Reassociation kinetics of *Lilium* DNA. Labeled DNA was prepared from microsporocytes that were cultured for two days in the presence of either ^3H -thymidine (10 $\mu\text{Ci/ml}$) or carrier-free ^{32}P -phosphate (10 $\mu\text{Ci/ml}$). Three groups of microsporocytes at different stages of meiosis were thus incubated: premeiotic S-phase,

—○—○—○—; zygotene —△—△—△—; pachytene —×—×—×—.

The procedures used in sonication and reassociation are described under Methods.

Reassociation of premeiotic S-phase DNA was measured by both UV absorbancy and radioactivity. Since the two curves coincided completely only one set of symbols is shown. *E. coli* DNA: —○—○—○—.

for C_{0t} values. In one procedure, a preparation of DNA from Z-labeled cells was reassociated to a C_{0t} value of 2.5×10^4 , the remainder concentrated, and its reassociation kinetics determined. In the second procedure, labeled Z-DNA was removed and pooled over a period of several months from a large number of CsCl gradients. The isopycnicly purified sample was also subjected to C_{0t} analysis. The results of these experiments are shown in Figure 2. The reassociation behavior of Z-DNA in the C_{0t} fractionated preparation is similar to that shown in Figure 1 and roughly approximates 2nd order kinetics. The optical density profile of that same preparation which includes all sequences of high C_{0t} value only partly tracks

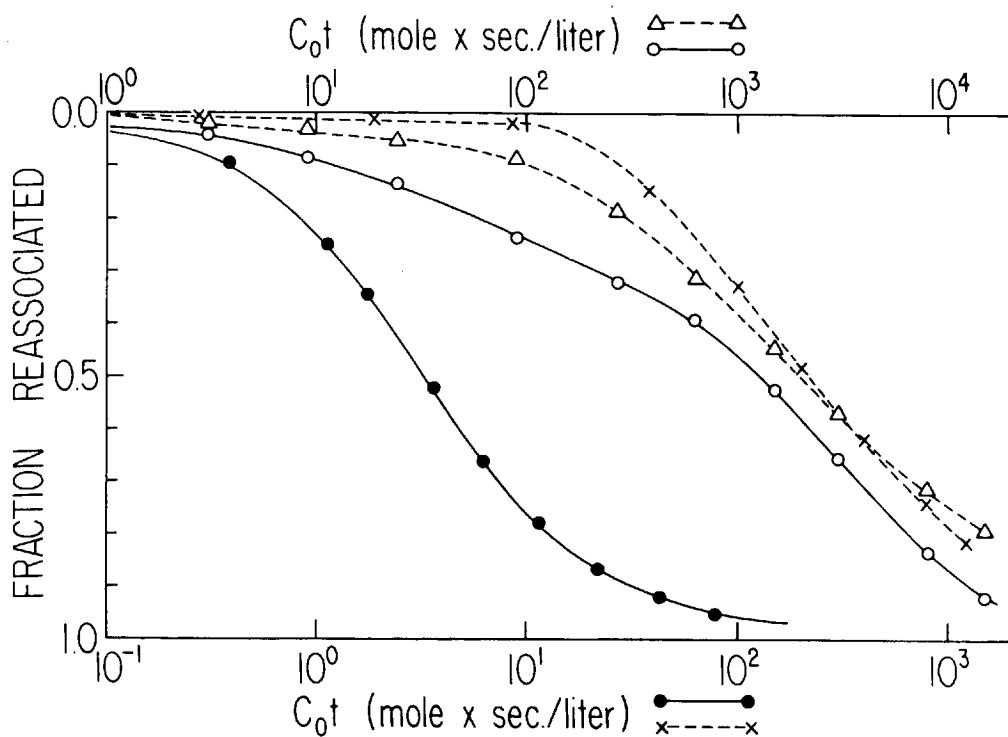


Figure 2. Reassociation kinetics of partially purified Z-DNA fractions. The denatured fraction remaining after reassociating zygotene labeled DNA to a C_{0t} value of 2.5×10^4 was concentrated by dialysis against 40% polyethylene glycol:0.12 M phosphate buffer. The solution was heated to 100°C to ensure denaturation and reassociation was measured in the standard way. Z-DNA radioactivity ---- Δ ---- Δ ----; optical density --- \circ --- \circ ---; Z-DNA purified by isopycnic centrifugation ---- \times ---- \times ----. The optical density and radioactivity profiles are coincident. *E. coli* DNA --- \circ --- \circ ---. The ^{14}C label represents Z-DNA. A: DNA from cells harvested at pachytene. B: DNA from cells harvested at the tetrad stage. Direction of gradient from right to left.

the Z-DNA radioactivity. Although the kinetics of the profile are difficult to analyze, it is probable that a significant concentration of unique sequences is present. Since the preparation contains about 30% of the total DNA, the unique portion of the optical density profile cannot be accounted for by the labeled Z-DNA. If Z-DNA consists of unique sequences, it can represent only a small fraction of the unique sequences in the *Lilium* genome. A much better separation of Z-DNA from unlabeled sequences was achieved by isopycnic purification (Figure 2). Unlike the C_{0t} purified Z-DNA, the radioactivity and absorbancy reassociation profiles are fully superimposed and provide a fairly good example of 2nd order reassociation. The similarity in reassociation kinetics of the two labeled preparations is nevertheless evident, and since they are products of two very different procedures, we conclude that 'satellite' Z-DNA is a collection of sequences that are either unique or in copies of equal numbers

The possibility that Z-DNA might consist almost entirely of single-copy sequences was examined by estimating its complexity. Z-DNA accounts for 0.3% of the lily genome or 1.8×10^{11} daltons ($6 \times 10^{13} \times 3 \times 10^{-3}$) of DNA (Hotta and Stern, 1974). If present solely as single-copy sequences, Z-DNA would have a complexity of 3×10^8 nucleotide pairs. The $C_{0t_{1/2}}$ value should therefore be 70 times that of *E. coli* DNA (Britten and Kohne, 1968). In Figure 1 the corrected $C_{0t_{1/2}}$ value of Z-DNA is estimated as 240 ($8 \times 10^4 \times 3 \times 10^{-3}$) and is thus 54 times ($240/4.5$) that of the internal *E. coli* standard. In Figure 2, the C_{0t} value of the isopycnicly purified Z-DNA may be read directly and is 66 times that of *E. coli*. The two ratios are essentially the same and are in the range expected for single-copy sequences.

About 30% of the total DNA remains unassociated at C_{0t} values above 2.5×10^4 (Figure 1). As a rough approximation, 60% of this fraction, which is analyzed in Figure 2, reassociates with 2nd order kinetics. If this fraction, which represents 18% of the genome, consists of unique sequences it should have 60 times the complexity of Z-DNA and a $C_{0t_{1/2}}$ value 4.2×10^3 times *E. coli* DNA. The ratio obtained from Figure 2 is 1.2×10^3 ($4.2 \times 10^3/3.5$). The disparity probably reflects errors in the estimates and/or a paucity of single-copy sequences. Each of these interpretations is compatible with the conclusion that Z-DNA represents single-copy sequences, and that about 18% of the genome consists of sequences repeated 1-10 times.

The precise distribution of Z-DNA within the chromosome is difficult to determine. It has been shown by autoradiography that Z-DNA label is distributed more or less uniformly over all *Lilium* chromosomes (Ito and Hotta, 1973). Such distribution, though determined imprecisely, is compatible with a synaptic role, but it does

not reconcile satellite behavior in CsCl gradients with the sequence characteristics observed in C_0t analyses. The apparent paradox was resolved by analyzing the sedimentation properties of Z-DNA at successive stages of meiosis.

Z-DNA behaves as a discrete class of GC-rich molecules only if prepared by conventional techniques and from cells that are in zygotene or pachytene. It does not behave this way if prepared from protoplasts by a much gentler procedure that minimizes shearing (Hotta and Stern, 1974). Under gentle conditions of extraction, Z-DNA prepared from prophase cells co-sediments with mainband DNA (Figure 3). A real, though labile, linkage appears to exist between Z-DNA and the genome as a whole such that shearing of the gently extracted DNA exposes a satellite profile. The labile linkage could be due to the presence of gaps at the ends of the strands being

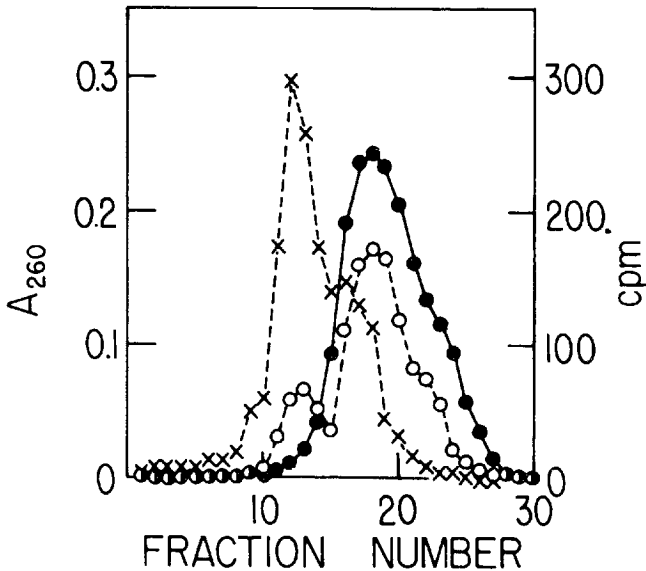


Figure 3. Isopycnic profiles of Z-DNA under different conditions of DNA isolation. Microsporocytes were cultured for two days *in vitro* in the presence of ^3H -thymidine (10 $\mu\text{Ci}/\text{ml}$). They were then washed and converted to protoplasts (Hotta and Stern, 1974). One aliquot of protoplasts was lysed in 1% sodium lauryl sulfate; 10 μM EDTA; 0.2% 'Brij' 58; 10 μM Tris buffer, pH 8.5. The lysate was very gently mixed with CsCl and then placed on the gradient $\text{---}\circ\text{---}\circ\text{---}$. The other aliquot of protoplasts was used to prepare DNA in the conventional way $\text{---}\times\text{---}\times\text{---}$. Both samples were centrifuged to equilibrium in a CsCl gradient. The optical density profile $\text{---}\circ\text{---}\circ\text{---}$ is the same regardless of method used. The small heavy peak in the gently prepared material is probably the result of handling. Stirring the latter preparation at 2,000 rpm for 5 min or heating to 80°C for 20 min with a little shaking substantially increases the size of the heavy peak.

synthesized in the late replicating Z-DNA regions. Their complementary strands, on the other hand, would presumably be integrated within the strands originally belonging to the parent DNA molecule. Zygotene regions in DNA duplexes prepared from zygotene cells might therefore be bracketed by single-strand nicks. The presence of such nicks could favor shearing into discrete Z-DNA molecules.

To test this explanation, microsporocytes were labeled during zygotene and then converted to protoplasts (Hotta and Stern, 1974). Both native and denatured DNA preparations were made from the protoplasts and their sedimentation characteristics determined. S-phase labeled cells were similarly processed for comparison. Results of

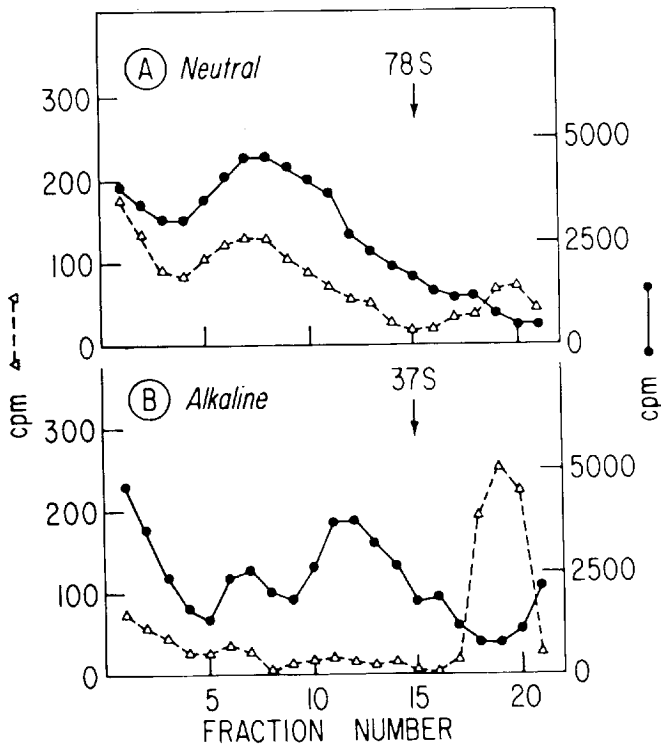


Figure 4. Sedimentation profiles of native and denatured Z-DNA. Meioocytes were labeled either during premeiotic interphase —○—○— for two days with ^3H -thymidine ($5\ \mu\text{Ci/ml}$) or during zygotene ---- Δ ---- Δ ---- for two days with ^3H -thymidine ($20\ \mu\text{Ci/ml}$). (A). Cells converted to protoplasts and lysed directly on gradient. The tubes were spun in an SW 27 rotor (Spinco) at 25 rpm for 8 hours (Hotta and Stern, 1974). (B). Cells were lysed in alkaline medium and centrifuged in an alkaline glycerol gradient. Direction of sedimentation from right to left.

these experiments are shown in Figure 4. The profile of the un-sheared native Z-DNA tracked that of S-phase labeled DNA as would be expected if Z-DNA were integrated with the remainder of the genome (Figure 4A). By contrast, denatured Z-DNA showed a sharp peak in the 10-12S region, whereas denatured S-phase DNA peaked at much higher S values (Figure 4B). The presence of a discrete class of low molecular weight single-strand Z-DNA molecules supports the speculation that the labile links between Z-DNA and the remainder of the genome are due to single-strand gaps at ends of the Z-DNA stretches.

The process of integrating zygotene with main band DNA was analyzed by comparing the sedimentation properties of labeled Z-DNA that was extracted under identical conditions at pachytene and at the termination of meiosis. If prepared by conventional techniques, the isopycnic behavior of Z-DNA isolated from pachytene cells differed

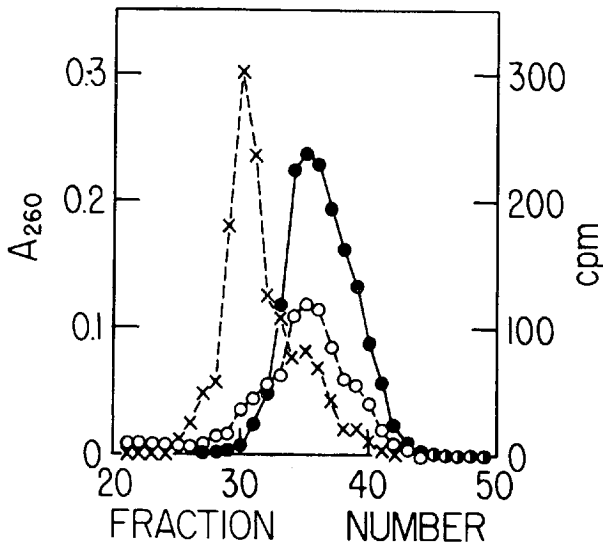


Figure 5. Integration of Z-DNA with mainband DNA at termination of meiosis. Microsporocytes were labeled during zygotene as in Figure 4. At the end of the two day incubation period, the cells were washed with medium and returned to culture in the presence of 0.2 mM cold thymidine. The culture was terminated either at the beginning of pachytene or at the tetrad stage that follows telophase II. DNA was isolated by conventional methods and the preparations centrifuged to equilibrium in a gradient of CsCl. Cells harvested at pachytene ----x----x----; cells harvested at tetrad stage----o----o----. The absorbancy curve ———o———o——— represents the bulk of the DNA and its profile remained unchanged in these different procedures.

from that of Z-DNA isolated at the termination of meiosis (Figure 5). In the latter case, Z-DNA radioactivity was distributed similarly to bulk DNA in the CsCl gradient. Shearing of such DNA did not reveal a typical Z-DNA satellite. The integration of Z-DNA into main band DNA must therefore have occurred at some time between pachytene and the termination of the second meiotic division.

Extraction of the labeled DNA under alkaline conditions provided confirmatory results. In this case, the preparations were characterized by their sedimentation rates. If extracted from pachytene cells, labeled Z-DNA showed, as already described, a peak in the 10-12S region. On the other hand, if extracted from cells that had largely completed meiosis, a low molecular weight component was no longer evident but, instead, the sedimentation profile of the Z-DNA mimicked that of bulk DNA (Figure 6).

As discussed earlier, the gaps bracketing Z-DNA regions should be present in those DNA strands that were newly synthesized during premeiotic S-phase. This supposition was tested by exposing S-phase microsporocytes to BdUrd. Premeiotic cells thus treated were cultured during zygotene in the presence of ^{14}C -thymidine. The DNA was then extracted either from cells that had reached pachytene or from those that had largely completed meiosis. The purified preparations were analyzed in alkaline CsCl gradients. The results are shown in Figure 7. Z-DNA isolated from pachytene cells sediments separately from the BdUrd-substituted strands. By contrast, Z-DNA isolated from second division cells behaves as though covalently linked

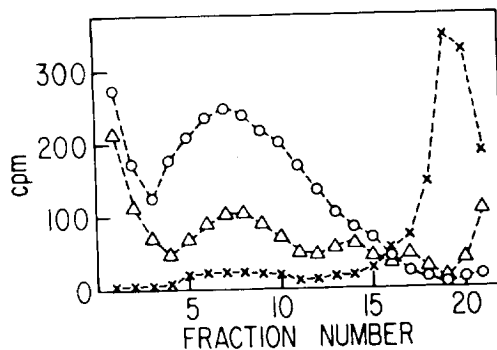


Figure 6. Sedimentation profiles of denatured Z-DNA during and after meiosis. Microsporocytes were labeled and cultured as in Figure 5. The cells were harvested in alkaline medium and the DNAs prepared from these were centrifuged in an alkaline glycerol gradient. Cells isolated at the pachytene stage ---x---x---; cells isolated at tetrad stage ---Δ---Δ---; curve for mainband DNA (data from Hotta and Stern, 1974) ---o---o---. Direction of sedimentation from right to left.

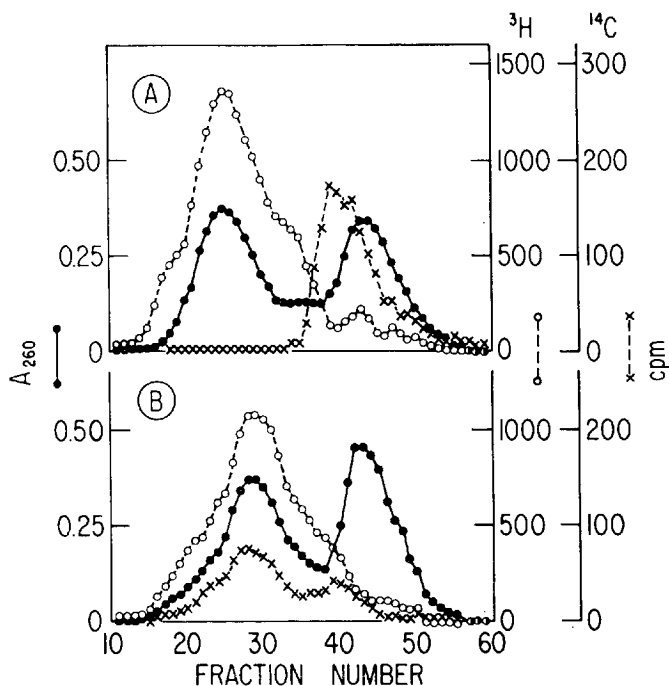


Figure 7. Isopycnic centrifugation of BU-substituted mainband DNA and Z-DNA under alkaline conditions. Excised flower buds at the premeiotic S-phase were cultured in the presence of 10 μ M BdUrd and 3 H-deoxyadenosine (15 μ Ci/ml) for three days. The microsporocytes were then extruded into culture medium containing 4 mM BdUrd for two hours and then transferred into 2 mM BdUrd; 3 H-deoxyadenosine (5 μ Ci/ml) for five days. At this stage most of the cells were in early zygotene and they were transferred to media containing 0.1 mM deoxyadenosine and 14 C-thymidine (10 μ Ci/ml) for three days, at which point half the cultures were harvested. The remainder was cultured for an additional 10–14 days so that most, but not all, cells reached the tetrad stage. The cells were converted to protoplasts and these were then lysed in 0.3N NaOH and centrifuged in an alkaline CsCl gradient. Absorbance —○—○—○—; cpm 3 H ---○---○---; cpm 14 C ----x----x----. The tritium label tracks the BU-substituted DNA and represents the strands synthesized during premeiotic S-phase. The 14 C-label represents Z-DNA. A: DNA from cells harvested at pachytene. B: DNA from cells harvested at the tetrad stage. Direction of gradient from right to left.

to the heavy DNA strand. All the experiments reported in Figures 4–7 support the conclusion that the lability of the Z-DNA linkage is the basis of its satellite behavior and that the lability itself is at least partly due to the presence of gaps in the newly synthesized DNA strands.

The delayed ligation of Z-DNA stretches to mainband DNA most probably reflects a particular meiotic function, possibly chiasma resolution. This function might be related to the suggested role

of Z-DNA regions as matching sites for homologous chromosomes. Since the matching function is presumed to be due to unique sequences, one way to examine such a role is by determining the degree of sequence homology between Z-DNA from different species of *Lilium*, particularly from those that do not form synaptic hybrids. If Z-DNA is the agent of synaptic matching, then parents of asynaptic hybrids should show less homology than those of synaptic ones in their Z-DNA component. Thus far we have tested only commercial hybrids of mixed parentage and one species of *L. speciosum*. The latter is probably not among the parents of the hybrids. Though inadequate, the comparisons are of interest (Table 1). The degree of sequence homology as measured by DNA-DNA interaction between cultivars of different hybrid lilies is of the order of 80% or higher when total nuclear DNA is tested. However, the level of sequence homology for Z-DNA falls between 24-40%. These results are consistent with the speculation that Z-DNA stretches may function in synapsis. However, regardless of the

TABLE 1
Sequence homology between different cultivars of *Lilium*

Fixed DNA	Hybridization of test DNA (% of control)						
	Fraction	BS	C	S	LW	SA	Average
C	Total	86		106	87	85	91
	Z	61	100	51	9	-	40
S	Total	88	95		100	109	93
	Z	40	64		6	26	34

Hybrid formation was determined by fixing a measured amount DNA to a nitrocellulose filter, usually 70-80 μg ('Fixed DNA') and incubating the filter in 6 x SSC at 68°C for 24 hours in the presence of about 500 $\mu\text{g}/\text{ml}$ of the DNA to be tested ('Test DNA') (Hotta and Stern, 1971). The values recorded are for saturation levels and these are expressed as a percentage of autologous values. Z-DNA was labeled in the standard way and was partially purified by collecting 10% of the total DNA from the heavy end of the CsCl gradient. The remaining 90% contained no radioactivity and was labeled with tritiated dimethylsulfate as described by Smith et al., 1967. This latter fraction is classified as 'Total DNA'. Labeled DNA was used in the test fractions and cold DNA, similarly prepared, was used in the fixed fractions. After hybridization, each filter was incubated for 10 min in a large volume of 1 x SSC at 50°C and then washed with 50 ml of 1 x SSC. The washed filters were dried and counted in the usual way.

BS = Bright Star; C = Cinnabar, S = Sonata; LW = *L. speciosum* var. Lucie Wilson; SA = *L. speciosum-auratum*.

validity of the speculation, which needs to be tested with appropriate interspecific comparisons, it is of interest that single-copy sequences should display so low a degree of homology between members of a genus that is karyotypically and, to a considerable extent, morphologically uniform.

Pachytene-labeled sequences It is generally recognized that the gross geometry of paired bivalents is inadequate to account for the intimate molecular pairing essential to crossing-over. Evidence has nevertheless been provided by a number of investigators that crossing-over does occur during pachytene (Henderson, 1970). Chromosome organization in the pachytene bivalent must therefore accommodate the needs of crossing-over, an accommodation that certainly includes the DNA moiety. Studies in our laboratory have provided several items of information concerning DNA behavior at pachytene (Stern and Hotta, 1973, 1974; Hotta and Stern, 1974; Smyth and Stern, 1973): (1) DNA synthesis during pachytene (P-DNA) is of the repair-replication type; (2) Repair occurs in co-ordination with the initiation of single-strand nicking at pachytene; (3) The sites of nicking are not spaced randomly in the DNA so that at pachytene, cells contain a complement of DNA fragments of relatively narrow size distribution; (4) Single-strand gaps disappear with the progress of P-DNA synthesis; (5) P-DNA synthesis occurs preferentially at sites containing repeated sequences. Pachytene spermatocytes of the rat have a DNA metabolism similar to that observed in *Lilium*, and P-DNA repeats are broadly distributed within the nuclear genome. Also, P-DNA sequences, though still undefined, are different from sequences present in the genome as a whole.

The C_{ot} curve of P-DNA from the Sonata cultivar shown in Figure 1 is similar to that previously reported by Smyth and Stern (1973) for *L. henryi*. Its profile is distinctly different from either total DNA or Z-DNA in that repeated sequences are comparatively abundant. If 0.1% is estimated as the fraction of DNA synthesized during pachytene, the synthesis accounts for 1×10^8 base pairs per haploid genome. If all pachytene sequences are labeled, the corrected $C_{ot}_{1/2}$ value should be 0.05 (50×10^{-3}).

Based on *E. coli* DNA, a $C_{ot}_{1/2}$ value of 0.05 signifies a complexity of only 5.0×10^4 base pairs. P-DNA sequences thus appear to be repeated 2×10^3 times per genome. However, all pachytene sites are probably not labeled but, as yet, we have no way of determining their proportion. Nevertheless, even though the estimates of the repeat number and complexity are imprecise, it is reasonable to conclude that P-DNA metabolism preferentially involves moderately repeated sequences.

The distinctiveness of *Lilium* P-DNA regions was also examined by determining the distribution of ^{32}P -label among pyrimidine tracts. Samples of S-phase labeled DNA were run in parallel for comparison. The low levels of pachytene labeling preclude, for the present at least, an absolute analysis of the concentration and composition of each of the tracts. A comparison of the relative distribution of label among S-phase and P-DNA tracts is nevertheless instructive (Table 2). Three differences have been consistently observed in the four analyses undertaken. The proportion of single pyrimidines is always higher in pachytene DNA, whereas the proportion of tetra-pyrimidines is always higher in the S-phase label. In P-DNA, specific activity falls off sharply in 7-pyrimidine and higher order tracts whereas, as expected, there is relatively little variation in S-phase DNA. There is thus a lower frequency of pyrimidine cluster-

TABLE 2
Relative specific activities of pyrimidine tracts
in S-phase and pachytene-labeled DNA

No. of residues	I	II	III	IV	V	VI	VII	VIII	SALT	ACID
S-phase	1.33	1.08	1.0	0.92	0.85	0.83	0.83	0.86	0.97	1.02
Pachytene	1.70	1.01	1.0	0.59	1.07	0.92	0.54	0.49	0.23	0.23

^{32}P -labeled pachytene DNA was obtained by culturing pachytene microsporocytes in the presence of ^{32}P -phosphate. Labeled S-phase DNA was obtained from very young anthers in buds of 2-5 mm length. Cut stems carrying a few such buds were kept in ^{32}P -phosphate solution (10 $\mu\text{Ci/ml}$). The leaves were removed from the stems and a stream of air was blown across the terminal flower buds for a period of 5 days. Under these conditions buds grew an additional 1-2 mm in length. The transpiration stream promoted the transport of radioactivity to the buds. DNA extracted from the entire anthers provided S-phase labeled DNA.

The purified DNA samples were hydrolyzed in 70% formic acid containing 2% diphenylamine in the dark for 18 hours at 30°C. At the end of the digestion period, the solution was shaken with ether and placed *in vacuo* to remove formic acid (Southern, 1970). After adjusting pH to 8.5 the digest was chromatographed on a DEAE-52 column. The procedure was the same as that used by Brunk (1973) except that elution was carried out at pH 7.5 instead of pH 5.5.

The specific activity of the trinucleotide (III) was set as 1 for each of the digests and the activities of the other residues are expressed relative to it. Direct comparison of radioactivities between the two samples is not possible since labeling conditions are necessarily different. 'Salt' and 'Acid' are the residual oligonucleotides eluted with 1 M NaCl and 0.1 N HCl respectively.

ing in P-DNA than in the genome as a whole. Pyrimidine tract analysis, and so too C_{0t} profiles, point to the conclusion that repair replication, which is presumably related to crossing-over, does not occur randomly along the DNA of the chromosomes.

The distribution of P-DNA repeats in CsCl gradients was analyzed to determine whether they were evenly represented in all the isopycnic fractions. Pachytene-labeled DNA was centrifuged to equilibrium as shown in Figure 8, and the gradient divided into three fractions for C_{0t} analysis. C_{0t} profiles of all three P-DNA fractions were essentially identical (Figure 9). Since the profile of radioactivity more or less tracked the absorbancy profile in the gradient (Figure 8), we conclude that the repeated sequences are evenly distributed among the three fractions.

An additional analysis was undertaken to determine whether a significant fraction of the P-DNA label was present in the high C_{0t} fraction. The unassociated strands remaining at C_{0t} values above 10^3 were pooled, concentrated, and their C_{0t} profile determined (Figure 9, curve 2). The presence of low repeats in the pachytene-labeled DNA is evident from the figure. The higher C_{0t} component could not have resulted from contamination by S-phase cells because such contamination is extremely rare for pachytene cultures and the pattern has been repeatedly observed. Neither could the component

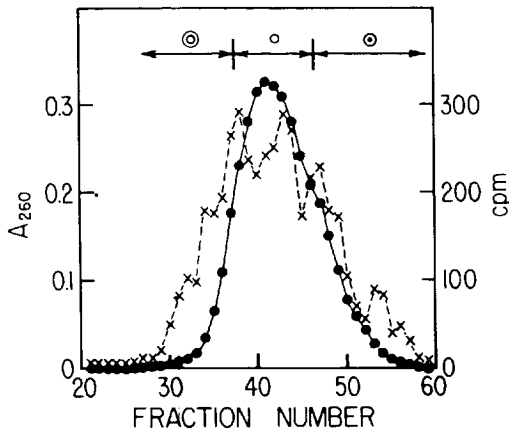


Figure 8. Radioactivity profile of pachytene-labeled DNA. Pachytene microsporocytes were incubated for two days in the presence of ^3H -thymidine. Procedures were otherwise the same as those used in labeling Z-DNA. The purified DNA was centrifuged to equilibrium in a CsCl gradient and the sample then divided into three fractions as shown by the arrows. Each fraction was concentrated by dialysis against 0.12 M phosphate buffer and polyethylene glycol. Radioactivity ----x-----; absorbancy at $260\text{m}\mu$ —○—○—. Direction of gradient from right to left.

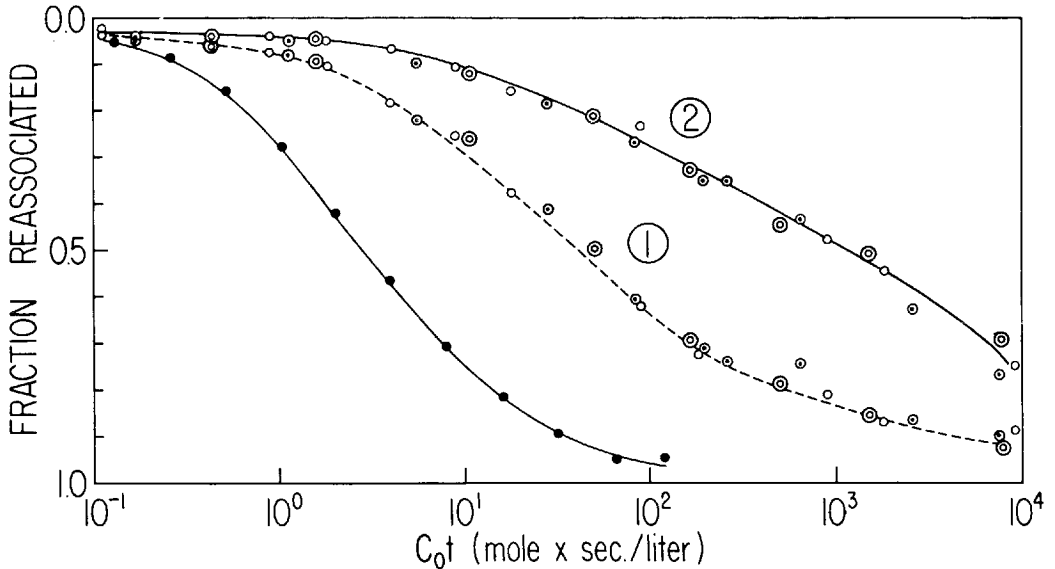


Figure 9. C_0t profiles of pachytene-labeled DNA. Three fractions of different mean buoyant density were obtained as described in Figure 8. Symbols used are explained in Figure 8. The profiles of the three fractions were essentially identical (curve 1). The unassociated strands with C_0t values greater than 10^3 were pooled and concentrated and a C_0t profile obtained (curve 2). In this case absorbancy and radioactivity coincided. The lower curve is *E. coli* DNA.

have been derived from zygotene cells because the reassociation kinetics are different from those observed in zygotene preparations. We conclude that the label reflects lower order repeats which are probably linked to the repetitive stretches.

The data obtained reinforce earlier observations that P-DNA metabolism, unlike repair-replication in general, occurs preferentially in regions of repeated sequences, and that these regions are widely distributed within the DNA of the genome. We do not know the length of P-DNA clusters and it may well be that their length varies from region to region. Neither do we know the diversity of the P-DNA regions within and between the chromosomes. Despite these major gaps in our information, there are sufficient grounds for concluding that dispersed regions of moderately repeated sequences have a special role in the processes associated with crossing-over.

DISCUSSION

We have reported on two different categories of DNA sequences that are metabolically active during meiosis. One consists of single-copy sequences, Z-DNA, and the other consists of moderately repeated sequences, P-DNA. Both are probably related to the processes of chromosome alignment and crossing-over, but the essentials

of the relationships are unknown and remain largely the subjects of speculation.

Our preferred function for Z-DNA is to align homologs in the course of synapsis by serving as matching sites for that alignment. The length of these zygotene sites may be estimated, albeit roughly, from sedimentation rates and is in the neighborhood of $3-4 \times 10^3$ base pairs. Interzygotene regions should therefore measure about 1×10^6 base pairs. A reasonable model of bivalent organization to accommodate Z-DNA function is one in which the Z-DNA stretches maintain a fixed and constant position during synapsis along the chromosome axis whereas the interzygotene stretches are variably positioned and may have short segments protruding into the synaptonemal complex (Stern and Hotta, 1974). These protruding stretches of DNA would be potential regions of crossing-over. The number of interzygotene stretches (calculated from the length of the zygotene region) are about $7-8 \times 10^3$ per chromatid. The experimentally observed spacing of nicks at pachytene is about 1.6×10^5 base pairs (Hotta and Stern, 1974), so that, if uniformly distributed, there would be about 7 nicks per interzygotene region. If bivalents are so structured that no more than one cross-over could occur in a single interzygotene region and none in the zygotene regions themselves, the distribution of Z-DNA stretches would define a subunit of positive interference. If synapsis does in fact depend upon the homology of Z-DNA regions, then from an evolutionary standpoint, Z-DNA would have an important role in speciation.

The size of P-DNA sequences may be estimated from the number of base pairs per genome (1×10^8) and the number of nicked sites (6×10^5) as determined by sedimentation analyses of DNA extracted from pachytene cells (Hotta and Stern, 1974). The value calculated is 170 base pairs per site, a relatively small size for a repeated sequence. These relatively short repeats probably occur within the interzygotene regions and we presume that they facilitate heteroduplex formation in association with crossing-over. Although we have very limited knowledge about their composition, it appears from pyrimidine tract analysis that they are probably not composed of a few simple sequences. It is possible that composition of the repeat regions is obscured by contamination from the less repeated sequences that are labeled during pachytene. Since we have no method as yet to resolve this particular question, we prefer to stay with the tentative conclusion that P-DNA consists of moderately repeated sequences that are distributed at more or less regular intervals along the chromosomes. We do not know the fraction of sequences labeled during any single pachytene interval, and so we cannot estimate the absolute number of pachytene repeats per genome, nor can we approximate the distances between them.

This work was supported by grants from the N.S.F. (GB 29562) and the N.I.H. (HD 03015).

REFERENCES

- Britten, R.J. and D.E. Kohne. 1968. Repeated sequences in DNA. *Science* 161: 529.
- Brunk, C.F. 1973. Distribution of dimers in ultraviolet-irradiated DNA. *Nature New Biol.* 241: 74.
- Gillies, C.B. 1973. Ultrastructural analysis of maize pachytene karyotypes by three dimensional reconstruction of the synaptonemal complexes. *Chromosoma* 43: 145.
- Henderson, S.A. 1970. The time and place of meiotic crossing-over. *Ann. Rev. Genet.* 4: 295.
- Hotta, Y. and A. Bassel. 1965. Molecular size and circularity of DNA in cells of mammals and higher plants. *Proc. Nat. Acad. Sci. USA* 53: 356.
- Hotta, Y., M. Ito and H. Stern. 1966. Synthesis of DNA during meiosis. *Proc. Nat. Acad. Sci. USA* 56: 1184.
- Hotta, Y. and H. Stern. 1971. Analysis of DNA synthesis during meiotic prophase in *Lilium*. *J. Mol. Biol.* 55: 337.
- _____, _____. 1974. DNA scission and repair during pachytene in *Lilium*. *Chromosoma*. In press.
- Ito, M. and Y. Hotta. 1973. Radioautography of incorporated ^3H -thymidine and its metabolism during meiotic prophase in microsporocytes of *Lilium*. *Chromosoma* 43: 391.
- Menzel, M.Y. and J.M. Price. 1966. Fine structure of synapsed chromosomes in *F₁ Lycopersicon esculentum-Solanum lycopersicoides* and its parents. *Amer. J. Botany* 53: 1079.
- Roth, T.F. and M. Ito. 1967. DNA-dependent formation of the synaptonemal complex at meiotic prophase. *J. Cell Biol.* 35: 247.
- Smith, K.D., J.L. Armstrong and B.J. McCarthy. 1967. The introduction of radioisotopes into RNA by methylation *in vitro*. *Biochim. Biophys. Acta* 142: 323.
- Smyth, D.R. and H. Stern. 1973. Repeated DNA synthesized during pachytene in *Lilium henryi*. *Nature New Biol.* 245: 94.
- Southern, E.M. 1970. Base sequence and evolution of guinea-pig α satellite DNA. *Nature* 227: 794.
- Stern, H. and Y. Hotta. 1970. Culture of meiotic cells for biochemical studies. In *Methods in Cell Physiology* (D.M. Prescott, ed.), Vol. IV, p.497. Academic Press, New York.
- _____, _____. 1973. Biochemical controls of meiosis. *Ann. Rev. Genet.* 7: 37.
- _____, _____. 1974. DNA metabolism during pachytene in relation to crossing-over. *Genetics*. In press.
- Yunis, J.J. and W.G. Yasmineh. 1971. Heterochromatin, satellite DNA, and cell function. *Science* 174: 1200.

Regulation of genetic recombination in *Neurospora crassa*

D. G. Catcheside

Higher organisms show less recombination per unit of DNA than do procaryotes. In *Neurospora crassa* restriction of recombination is due to dominant regulatory genes (*rec*⁺) at three known *recombination* loci. Their products interact specifically with recognition genes at *control* (*con*) loci, usually not linked to the *rec* loci. Crossing over is reduced by a given *rec*⁺ around each related *con* locus and conversion in a neighbouring locus is also greatly diminished. Relaxation of recombination is due to dominant *cog*⁺ genes situated in target regions and expressed in the absence of the *rec*⁺ gene specific to the particular target. Relaxation is expressed strictly locally as increased crossing over and increased conversion, especially in the *cog*⁺ chromosome in a heterozygote. The number of regulatory *rec* loci is about ten and of *con* loci about 500 belonging to about ten species. The *cog* loci, believed to be targets of an endonuclease initiating recombination, probably also number about 500.

Compared with procaryotes, higher organisms show much less recombination per unit of DNA in the haploid set (Table 1). Since the mechanism of recombination is fundamentally the same in all organisms, involving basically similar enzymic mechanisms acting on similar substrates, this reduction must mean the acquisition of control in eucaryotes. Moreover, even within the same organism, different regions of the chromosomes of roughly equal dimensions show marked differences of frequency. For example, in yeast the frequency of conversion, one manifestation of recombination, varies from 0.3% at the mating type locus to 20% at the *sup 6* locus (Fogel and Mortimer, 1971). Conceivably control might be achieved in a number of ways.

It is observed in *Neurospora*, and the same appears to be true of other organisms, that the chromosomes are considerably condensed before pairing of homologues occurs (McClintock, 1945). A total gene by gene matching at zygotene therefore appears to be excluded and one may infer that intimate synapsis is restricted to special regions of the chromosomes and that these are the sites of initiation of recombination. These pairing regions might either remain exposed in the partially condensed chromosomes or be exposed by a specific mechanism. In either case the pairing regions must have special properties residing in the sequence of nucleotides in

their DNA. However, these sequences need not all be exactly alike.

TABLE 1

Relation of recombination frequency to DNA content in various organisms

Species	Map length	Nucleotide pairs in DNA	
		Per haploid nucleus	Per map unit
T4	800	2×10^5	2.5×10^2
<i>E. coli</i>	2000	10^7	5×10^3
Yeast	3700	2.2×10^7	6×10^3
<i>Neurospora</i>	500	4.3×10^7	8×10^4
<i>Drosophila</i>	287	2×10^8	7×10^5

A second level of control may exist at the stage at which a specific endonuclease initiates the events of recombination by nicking at a pairing region one or more chains of the pairs of DNA molecules which have synapsed. The pairing and initiation segments may differ from one another so that there is a greater or lesser probability of nicking according to how well the endonuclease recognises the specific sequence. With respect to the endonuclease the initiation segment behaves like a promoter. Evidence of a variant promoter is believed to exist at the *cog* locus near the *his-3* locus in *Neurospora crassa* (Angel et al., 1970). A case in *Schizosaccharomyces pombe* interpreted as a marker effect is probably similar (Gutz, 1971).

A third level of control has been found in *Neurospora crassa* involving the action of genes at three known loci: *rec-1*, *rec-2* and *rec-3*. These act specifically at selected places scattered in the chromosomes. At these places there must be specific recognition sites for the action of each gene.

Regulatory recombination (*rec*) genes and
recognition control (*con*) genes

Two alleles are known at each of *rec-1* (Jessop and Catcheside,

1965) and *rec-2* (Smith, 1966; Jha, 1967; Catcheside and Austin, 1969; Catcheside and Corcoran, 1973), while three alleles are known at the *rec-3* locus (Catcheside and Austin, 1971; Catcheside, 1974). In each case, one allele acts dominantly to depress recombination in particular targets, one dose of the dominant allele being as effective as two. The genes at each *rec* locus are specific to a few known target regions (Table 2), though there must be many more. The targets are usually not close to the *rec* locus and are often on different chromosomes (Figure 1). The distribution of *rec* genes in laboratory stocks of wild or derivative status is shown in Figure 2.

TABLE 2

Targets of products of regulatory *rec* genes in
Neurospora crassa

Locus	<i>rec-1</i>	<i>rec-2</i>	<i>rec-3</i>
<i>his-1</i>	+	0	0
<i>nit-2</i>	+	0	0
<i>his-3</i>	0	+	0
<i>am-1</i>	0	0	+
<i>his-2</i>	0	0	+
Others	0(14)*	0(11)*	0(7)*
Region			
<i>pyr-3 his-5</i>	0	+	0
<i>his-3 ad-3</i>	0	+	0
<i>arg-3 sn</i>	0	+	0
<i>sn his-2</i>	0	0	+

* Bracketed are numbers of loci unaffected by *rec* genes.

It seems probable that each *rec* gene controls allelic recombination at a number of loci and also crossing over in the neighbourhood of each target locus. This is definitely the case with the following: (1) *rec-3*⁺ reduces allelic recombination in the *his-2* locus and crossing over between *sn* and *his-2*; (2) *rec-2*⁺ reduces allelic recombination in the *his-3* locus and crossing over between

his-3 and *ad-3* (Angel et al., 1970); (3) *rec-1*⁺ reduces allelic recombination in the *nit-2* locus (D.E.A. Catcheside, 1970;1974) and crossing over around it. The effects of *rec-1*⁺ and *rec-3*⁺ respectively in reducing crossing over near to the *his-1* (Jessop and Catcheside, 1965) and *am-1* (Smyth, 1973) loci are not demonstrable. Probably the effects are within the limits of experimental error.

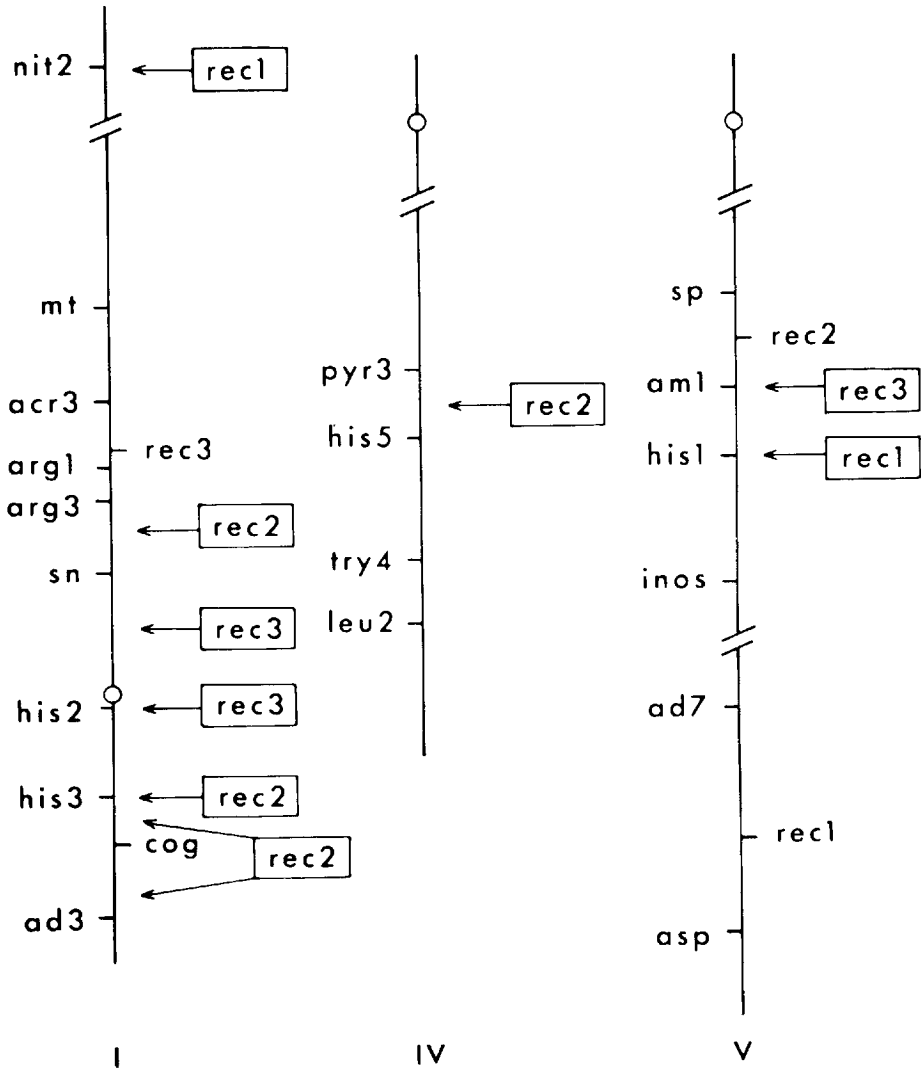


Figure 1. Maps of linkage groups I, IV and V of *Neurospora crassa* showing positions of *rec* and *cog* loci and the targets of their actions. The latter are indicated by boxed *rec* symbols attached to arrows.

The effects of *rec-2*⁺ in reducing allelic recombination in one of the boundary loci of the *pyr-3 his-5* and *arg-3 sn* segments cannot be examined except in the case of *his-5*. There are no effects on

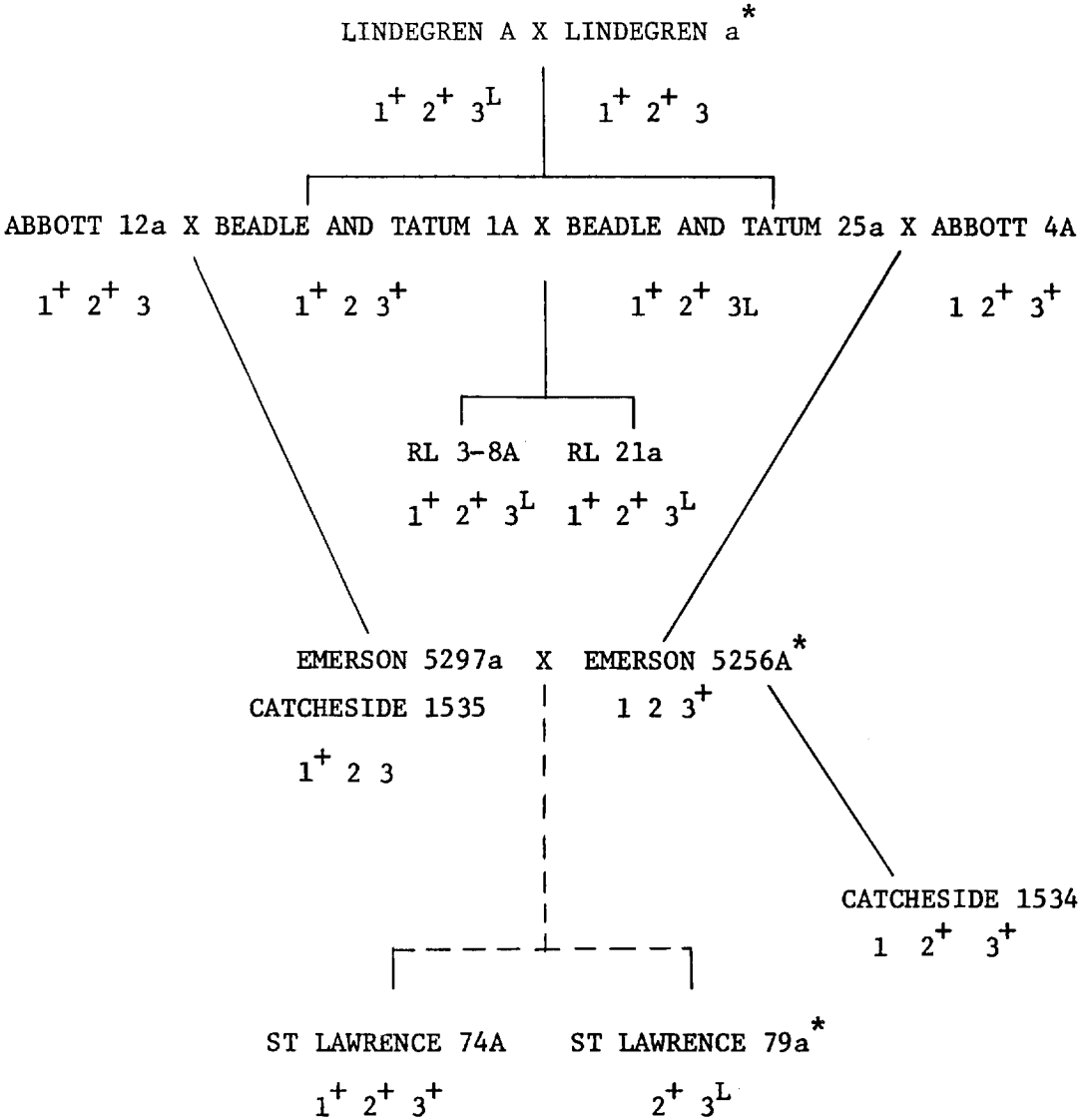


Figure 2. Pedigree of laboratory stocks of *Neurospora crassa* with their *rec* gene constitutions. The stocks marked with an asterisk are not descended directly from the originals and their stated *rec* constitutions are probably not those of the original stocks of the particular designation.

the frequency of allelic recombination at the *his-5* locus (Smith, 1966). Allelic crosses between mutants at the *pyr-3* and *arg-3* loci are virtually sterile and at the *sn* locus there is only one mutant gene.

Besides reducing the frequency of allelic recombination at a locus, a dominant *rec*⁺ gene also changes the polarity of conversion in the locus. This is manifested in changes in the relative frequencies of the four combinations of flanking markers amongst prototrophic recombinants formed at a locus. In a cross of the type $P m^1 + D \times p + m^2 \bar{d}, m^1$ and m^2 represent two mutant sites of difference at a locus with m^1 proximal to m^2 , the flanking markers P and p being proximal to the m locus and the markers D and d distal to it. A predominance of $P D$ over $p d$ among m^+ prototrophs indicates a higher frequency of conversion of m^1 and a direction of conversion originating proximally to the m locus. Thomas and Catcheside (1969) showed that besides a fifteen-fold reduction in the frequency of prototrophs at the *his-1* locus, the *rec-1*⁺ gene changed the ratio of $P D : p d$ from 1.5:1 in *rec-1* x *rec-1* to 0.55:1 in *rec-1*⁺ x *rec-1* and *rec-1*⁺ x *rec-1*⁺. This demonstrates that *rec-1*⁺ has an action on the *his-1* locus originating from near the proximal end of that locus.

Any given target region appears to be affected only by the genes at one of the *rec* loci and not by those at the other *rec* loci. This means that the target regions must have recognition sites (Catcheside, 1966) consisting of a sequence of DNA which is able to interact with the products of genes from one *rec* locus but not with those from the other *rec* loci. These recognition site genes may be referred to as *control* (*con*) genes. There would be a species of *con* gene, each of which could be called *con-1*, *con-2* and *con-3*, corresponding to each species of *rec* locus and there would be a variety of each species of *con* gene situated in each target region.

The probable position of the *con* locus can be deduced for most of the target regions on one or both of two criteria. If the frequency of crossing over is controlled in a segment but not outside it, there must be a *con* locus in the segment. If the frequency of allelic recombination is controlled in a target locus, the *con* locus would be near to the end of the target locus where the highest frequency of conversion is seen in relaxed (*rec*) crosses. These criteria would place a *con-1* locus proximal to *his-1*, *con-2* loci between *pyr-3* and *his-5* and between *his-3* and *ad-3* and *con-3* loci proximal to *am-1* (Smyth, 1971) and between *sn* and *his-2*. There would also be a *con-2* locus between *arg-3* and *sn* or even to the left of this segment. A *con-1* locus would also be near to *nit-2*.

At the present time there is no independent evidence of *con* loci, in the sense that no variant at a *con* locus has been observed

nor even sought in one of the target regions. The question of variation within a species of *con* gene will be considered later.

The degree of effect of *rec* genes varies considerably as between different target regions (Table 3). Thus the ratio of effect of *rec-2* versus *rec-2*⁺ ranges from two in the *arg-3 sn* segment to 12.6 in the *pyr-3 his-5* segment. The latter case is worth more detailed analysis as indicative of the magnitude of various effects. In *rec-2* x *rec-2* crosses, crossing over between *pyr-3* and *his-5* averages about 23.3% compared with 1.84% in *rec-2*⁺ x *rec-2* crosses. No differential effect appears to extend proximally

TABLE 3

Action of *rec* genes in *Neurospora crassa*

Place of action	<i>rec</i> gene	Crossed to		Ratio
		<i>rec</i>	<i>rec</i> ⁺	
(a) Allelic recombination*				
<i>his-1</i>	<i>rec-1</i>	43	3	14
<i>nit-2</i>	<i>rec-1</i>	47	10	5
<i>his-3</i>	<i>rec-2 cog</i>	20	4	5
	<i>rec-2 cog</i> ⁺	165	5	33
<i>am-1</i>	<i>rec-3</i>	25	1	25
	<i>rec-3</i>	8	1	8
<i>his-2</i>	<i>rec-3</i>	42	7	6
	<i>rec-3</i>	7	7	1
(b) Crossing over [†]				
<i>pyr-3/his-5</i>	<i>rec-2</i>	19	1	16
<i>his-3/ad-3</i>	<i>rec-2 cog</i>	2.6	2.1	1.2
	<i>rec-2 cog</i> ⁺	9	2	4.5
<i>sn/his-2</i>	<i>rec-3</i>	12	0.6	20

* Prototrophs per 10⁵. Allelic markers: *his-1*: K83, K625; *nit-2*: MN72, MN73; *his-3*: K504, K874; *am-1*: K314, 47305; *his-2*: K584, K612.

† Recombinants per cent.

to *pyr-3* nor distally to *his-5*. The absence of *rec-2*⁺ from a cross adds some 21.5 centimorgans to the genetic length of the *pyr-3 his-5* segment, which has a basic value of 1.8. Each meiotic cell in which a crossing over, presumably reciprocal, occurs between *pyr-3* and *his-5* would produce an equal number of recombinant and non-recombinant spores. Therefore the observed increase means the occurrence of a chiasma between *pyr-3* and *his-5* in 47% of meiotic cells in *rec-2* x *rec-2* compared with 3.7% in *rec-2*⁺ x *rec-2*. Work in recent years has tended to show that allelic (conversion) and non-allelic (crossing over) recombinations are different manifestations of one common mechanism. Moreover, half of the tetrads in which conversion has occurred show correlated crossing over between flanking markers, while the other half do not. We may infer that a precursory event, which could lead to conversion, as well as to crossing over in half of the affected tetrads, occurs between *pyr-3* and *his-5* in 94% of all meiotic cells in *rec-2* x *rec-2*. Recombination is virtually uncontrolled in this region in these crosses. There is still distinct control in the other segments affected by *rec-2* genes. A similar argument leads to primary effects in 26% of meiotic cells in the *his-3 ad-3* segment and to 12% in the *arg-3 sn* segment. Collectively *rec-2*⁺ removes an average of about half a chiasma in respect of these three segments. The observed differences in response of different regions suggest that the varieties of *con-2* in the three segments may not all be alike. However, other factors could bring about the different responses.

Evidence for varietal differences of *con-3*, the recognition genes for the products of genes at the *rec-3* locus, is provided in another way. At least three different alleles occur at the *rec-3* locus in the left arm of linkage group I. They are *rec-3* present in Emerson a, *rec-3*⁺ present in Emerson A and *rec-3*^L present in Lindegren A. The genes are also present in ancestors and derivatives of these wild strains as well as in many mutants derived from them (Figure 2). The three *rec-3* alleles differ in their action on allelic recombination at the *am-1* and *his-2* loci. The relative effects of *rec-3*⁺, *rec-3*^L and *rec-3* on allelic recombinations are 1 : 8 : 25 at the *am-1* locus and 7 : 7 : 42 at the *his-2* locus. Note that in every case the gene giving lower yields is dominant to that giving higher yields of recombinants. The *con-2* gene near to *his-2* evidently does not differentiate between the products of *rec-3*⁺ and *rec-3*^L respectively and must therefore be different in some minor respect from the *con-2* gene near to *am-1*. Non-allelic recombination between *sn* and *his-2* is apparently similar in *rec-3*^L crosses to the values in *rec-3*⁺ crosses.

The failure of the *con-2* gene near to *his-2* to differentiate between *rec-3*⁺ and *rec-3*^L raises doubts about the interpretation of

negative tests for the possible action of genes at the *rec-1*, *rec-2* and *rec-3* loci on a number of loci (Table 2). It is possible that the recessive alleles at each *rec* locus produce products of lower affinity for *con* genes and that the absence of a difference due to the dominant versus recessive *rec* genes reflects a failure of differentiation between their products by some *con* loci.

The possible dominance relations of variants at a *con* locus are difficult to predict. It is assumed that the interaction between a *rec* gene and its corresponding *con* genes occurs through affinity between a product of the *rec* gene and each of its corresponding *con* genes. There is a close analogy between regulatory gene and operator gene. The association of *rec* gene product and *con* gene might act (a) to frustrate pairing in the neighbourhood of the *con* locus or (b) to frustrate initiation of an act of recombination in the neighbourhood of the *con* locus after pairing. In the latter case the complex of *rec* product and *con* on one chromosome might act to prevent the initiation of recombination (nicking of a DNA chain) either (i) in both homologues or (ii) only in the homologue with the complex. Thus in a heterozygote for two allelic *con* genes having different affinities for the same *rec* product, cases (a) and (bi) would predict that low frequency of recombination, due to variation in *con* alleles, would be dominant to high frequency. Conversely, case (bii) would predict that low frequency would be recessive to high frequency and that initiation and conversion would be restricted very largely to one chromosome, that without or with a lower frequency of the complex. The only satisfactory evidence would come from a case of allelic *con* genes which responded differently to the same *rec*⁺ gene, since only in this case can we be sure that a regulatory substance is produced.

Recognition (*cog*) genes

The *cog*⁺ versus *cog* gene system (Angel et al., 1970) between *his-3* and *ad-3* could be interpreted as a case (bii), were it not for the different response being shown only in the absence of *rec-2*⁺. The *cog* locus is situated between *his-3* and *ad-3* about a quarter of the genetic distance from *his-3* to *ad-3*. The presence of *cog*⁺, if *rec-2*⁺ is absent from a cross, leads to a six-fold increase in the frequency of allelic recombination in the *his-3* locus. It also leads to a four-fold increase in crossing over between *his-3* and *ad-3*, the increase being equally distributed between the *his-3 cog* and *cog ad-3* segments. Apparently *cog*⁺ is completely dominant, for no differences in frequency are seen between *cog*⁺ x *cog* and *cog*⁺ x *cog*⁺. However, there are differences in other respects. In the *cog*⁺ x *cog* heterozygote the *his-3* site in the *cog*⁺ chromosome undergoes preferential conversion irrespective of whether it is proximal or distal to the *his-3* site in the homologous chromosome.

In cog^+ x cog^+ and cog x cog crosses the distal site is preferentially converted. This behaviour suggests that cog is the locus at which, after synapsis, the formation of a heteroduplex DNA is initiated, following nicking by an endonuclease. There is a strong preference for cog^+ over cog in the heterozygote and probably a heteroduplex is formed only on one chromatid of the tetrad (Catcheside and Angel, 1974). The cog locus appears not to be the target of the product of $rec-2^+$ (or of its allele $rec-2$). Hence the target of the $rec-2^+$ product must be yet another locus in the $his-3$ $ad-3$ region. The relation between the con and cog loci would be analogous to that between operator and promoter.

To interpret these cog genes as con genes, the targets also of rec gene products, it would be necessary to assume (1) that cog and cog^+ have similar affinities for the product of $rec-2^+$; (2) that $rec-2$ has a product; (3) that cog^+ has a lower affinity than cog for the product of $rec-2$ and (4) that the endonuclease nicks cog much less efficiently than it does cog^+ . Consequently, in the presence of $rec-2$, more recombination could be initiated in the cog^+ chromosome than in the homologous cog chromosome. Evidently more study is required of variant control genes in or adjacent to target segments of loci.

Numbers of recombination and control genes

The data available for making estimates of the numbers of rec and con genes are limited, but serve to indicate orders of magnitude. A fair number of loci have been tested for effects of regulatory genes at the $rec-1$, $rec-2$ and $rec-3$ loci. Five loci out of a total of about fifteen which have been satisfactorily assayed show regulation by genes at one or other of the three regulatory rec loci. Simple division suggests a total of nine rec loci, so the order of magnitude can be said to be ten.

As regards con loci, for which there would be several corresponding to each rec locus, an estimate can be derived in a different way. The total genetic map length in *Neurospora crassa* is estimated to be about 500 centimorgans. This is based on restricted values of crossing over, as judged from comparison between relaxed (rec) and restricted (rec^+) values in our data and corresponding values in published data. For the six segments for which regulation has been studied the total restricted length is 6 centimorgans, that is an average of one centimorgan per segment. On this basis there would be 500 segments each with a con locus in *Neurospora crassa*, an average of 50 con loci of a given species corresponding to each rec locus. If the cog loci are distinct from the con loci, there should be about 500 cog loci. These species of repetitious DNA would be scattered throughout the chromosomes.

REFERENCES

- Angel, T., B. Austin and D.G. Catcheside. 1970. Regulation of recombination at the *his-3* locus in *Neurospora crassa*. *Aust. J. Biol. Sci.* 23: 1229.
- Catcheside, D.E.A. 1970. Control of recombination within the *nitrate-2* locus of *Neurospora crassa*: an unlinked dominant gene which reduces prototroph yields. *Aust. J. Biol. Sci.* 23: 855.
- . 1974. A second locus subject to recombination control by the *rec-1*⁺ gene in *Neurospora crassa*. *Aust. J. Biol. Sci.* 27. In press.
- Catcheside, D.G. 1966. A second gene controlling allelic recombination in *Neurospora crassa*. *Aust. J. Biol. Sci.* 19: 1039.
- . 1974. The occurrence in wild strains of *Neurospora crassa* of genes controlling genetic recombination. *Aust. J. Biol. Sci.* 27. In press.
- Catcheside, D.G. and T. Angel. 1974. A *histidine-3* mutant, in *Neurospora crassa*, due to an interchange. *Aust. J. Biol. Sci.* 27: 219.
- Catcheside, D.G. and B. Austin. 1969. The control of allelic recombination at *histidine* loci in *Neurospora crassa*. *Am. J. Bot.* 56: 685.
- , ———. 1971. Common regulation of recombination at the *amination-1* and *histidine-2* loci in *Neurospora crassa*. *Aust. J. Biol. Sci.* 24: 107.
- Catcheside, D.G. and D. Corcoran. 1973. Control of non-allelic recombination in *Neurospora crassa*. *Aust. J. Biol. Sci.* 26: 1337.
- Fogel, S. and R.K. Mortimer. 1971. Recombination in yeast. *Ann. Rev. Genetics* 5: 219.
- Gutz, H. 1971. Site specific induction of gene conversion in *Schizosaccharomyces pombe*. *Genetics* 69: 317.
- Jessop, A.P. and D.G. Catcheside. 1965. Interallelic recombination at the *his-1* locus in *Neurospora crassa* and its genetic control. *Heredity* 20: 237.
- Jha, K.K. 1967. Genetic control of allelic recombination at the *histidine-3* locus of *Neurospora crassa*. *Genetics* 57: 865.
- McClintock, B. 1945. *Neurospora*. I. Preliminary observations of the chromosomes of *Neurospora crassa*. *Am. J. Bot.* 32: 671.
- Smith, B.R. 1966. Genetic controls of recombination. I. The *recombination-2* gene of *Neurospora crassa*. *Heredity* 21: 481.
- Smyth, D.R. 1971. Effect of *rec-3* on polarity of recombination in the *amination-1* locus of *Neurospora crassa*. *Aust. J. Biol. Sci.* 24: 97.

- Smyth, D.R. 1973. Action of *rec-3* on recombination near the *amination-1* locus of *Neurospora crassa*. *Aust. J. Biol. Sci.* 26: 439.
- Thomas, P.L. and D.G. Catcheside. 1969. Genetic control of flanking marker behaviour in an allelic cross of *Neurospora crassa*. *Canad. J. Genet. Cytol.* 11: 558.

The synaptonemal complex in the course of meiotic chromosome pairing and recombination

C. B. Gillies

Reconstructed serial electron micrographs from prophase I meiocytes have been used to study the relationship of the synaptonemal complex to synapsis and crossing over. At leptotene, localization of lateral component ends on the nuclear envelope near the nucleolus forms a type of bouquet arrangement. This may be a form of homologous chromosome alignment which results in zygotene initiation of synaptonemal complex formation near the ends of bivalents. Centromeres, knobs, and the nucleolus organizer regions do not seem to be involved in initiation of synapsis. The occurrence of synaptonemal complexes in structural heterozygotes and nonhomologously paired regions suggests that crossing over requires a final pairing step after the formation of the synaptonemal complex. This could involve DNA base pairing. The synaptonemal complex may, however, have a final role to play by stabilizing chiasmata in diplotene.

The mechanism and control of genetic recombination in eukaryotic chromosomes remains one of the central problems of cytogenetics. While the previous speakers have addressed these topics from biochemical and genetical viewpoints, it is the intent of this paper to examine some of the cytological mechanisms which seem to be intimately related to meiotic recombination.

It is fairly well established that crossing over occurs at the pachytene stage of meiosis (Lu, 1970), and that the presence of synaptonemal complexes between paired homologous chromosomes is a prerequisite for recombination (Westergaard and von Wettstein, 1972). However, the exact role of the synaptonemal complex in mediating crossing over is as yet uncertain. Likewise, the sequence of events which leads up to homologous chromosome pairing with fully formed synaptonemal complexes has not been completely determined. In the present study serial electron micrographs of meiotic stages have been reconstructed in order to study questions concerning the assembly and role of the synaptonemal complex.

METHODS

Prophase I stages of *Neurospora*, maize and haploid barley have been examined. The strains used in each case have been reported

previously (Gillies, 1972, 1973a, 1974). In maize and barley, anthers of appropriate meiotic stages were selected after light microscopic examination of adjacent anthers. In *Neurospora* an isolated ascus technique was used (Westergaard and von Wettstein, 1970a). Preparation for electron microscopy was by established techniques (Gillies, 1972, 1973b). Serial electron micrographs were prepared and reconstructed as previously described (Gillies, 1972).

RESULTS AND DISCUSSION

Synaptonemal complex and chromosome synapsis In maize at leptotene distinct lateral components can be seen (Figure 1), and they appear to be continuous along the length of the chromosome. Chromatin is relatively uncondensed, but chromatic knobs and centromeres are recognisable. In serial sections of one nucleus it was possible to discern 26 lateral component ends which terminated at the nuclear envelope. However, individual chromosomes could not be identified. These 26 ends were found to be grouped on a small area of the nuclear envelope close to the region of appression of the nucleolus (Figure 4). In some cases there was indication of grouping in pairs (Figure 4). The area of the nuclear envelope occupied by the nucleolus and the chromosome ends represented approximately one eighth of the nuclear envelope surface. This region appeared to have few nuclear pores.

At zygotene stages in maize, both completed stretches of synaptonemal complex and unpaired lateral components were present (Figures 2 and 3). Condensation of chromatin was more advanced than at leptotene, and knobs and centromeres were readily distinguishable. In three reconstructed zygotene nuclei approximately 30% of the lateral components were paired to form synaptonemal complexes. There were approximately 40 discrete sites per nucleus where synaptonemal complex formation had been initiated. Pairing of the lateral components occurred at several places in each bivalent (Figures 6 and 7), but one place was usually at or close to the nuclear envelope (Figures 3 and 6). Of 13 bivalent ends examined in one cell, 9 ended at the nuclear envelope in fully paired synaptonemal complexes. The other four had synaptonemal complexes near the nuclear envelope but their lateral components were unpaired where they attached to the nuclear envelope (Figure 5).

In all three reconstructed zygotene nuclei the identified bivalent ends were found clustered on the nuclear envelope near to the nucleolus (Figure 5). The area they occupied was only about 5% of the surface area of the nuclear envelope. They appeared more closely grouped than at leptotene.

At both leptotene and zygotene the centromeres of most chromosomes were recognisable. The frequency of fusion of the centromeric chromatin and the occurrence of paired synaptonemal complexes in centromeres was not high in either zygotene or leptotene. In the leptotene cell reconstructed, four homologous centromeres had fused (Figure 8). In the zygotene cells up to four cases were present of homologous centromeres fusing without their lateral components having paired. Two of the nuclei had single bivalents in which a synaptonemal complex passed through the centromere. Some obviously homologous centromeres were not fused (Figure 7). Similar observations were made of the associations of chromatic knobs at leptotene and zygotene. No obvious tendency to precocious association was seen at leptotene (Figure 8). At zygotene, some chromatic knobs had paired and synaptonemal complexes were present, but others showed completely unpaired lateral components.

The single nucleolus of a leptotene nucleus is attached to two homologous nucleolus organizing chromosomes at their nucleolus organizer regions (NOR). The NORs are ramified throughout the nucleolus in early leptotene, but contract to the periphery and form a single NOR by mid zygotene. The lateral components can be traced into and out of the NOR at leptotene (Figure 8). At zygotene, the two lateral components are located adjacent to each other in the single NOR, but in none of the three reconstructed zygotene nuclei was there synaptonemal complex formed in the NOR (Figure 7). In one case synaptonemal complex was present in the short arm of the nucleolus bivalent.

At pachytene chromosome ends were found attached to the nuclear envelope in *Neurospora*, maize and haploid barley. In haploid barley the nonhomologously paired chromosomes had a tendency to be matched at their ends, and cases of synaptonemal complex formation right up to the points of nuclear envelope attachment were noted (Figure 12).

The above results together with other published data suggest the following scheme of chromosome pairing in prophase I of plants: at *leptotene* the lateral components are synthesised on partly condensed chromosomes, incorporating only a small part of the chromatin but in a format which facilitates matching of homologous sites. Incorporated segments may be late replicating (Stern and Hotta, 1973). Some alignment of ends occurs, probably via their attachment to the nuclear envelope. A specialised portion of the nuclear envelope may be involved, possibly with the nucleolus having a function, hence the bouquet arrangement. At *zygotene* synaptonemal complex formation is initiated when lateral components approach closer than a critical distance (3000Å? Moens, 1969). This process is illustrated in Figure 3. The alignment of ends at

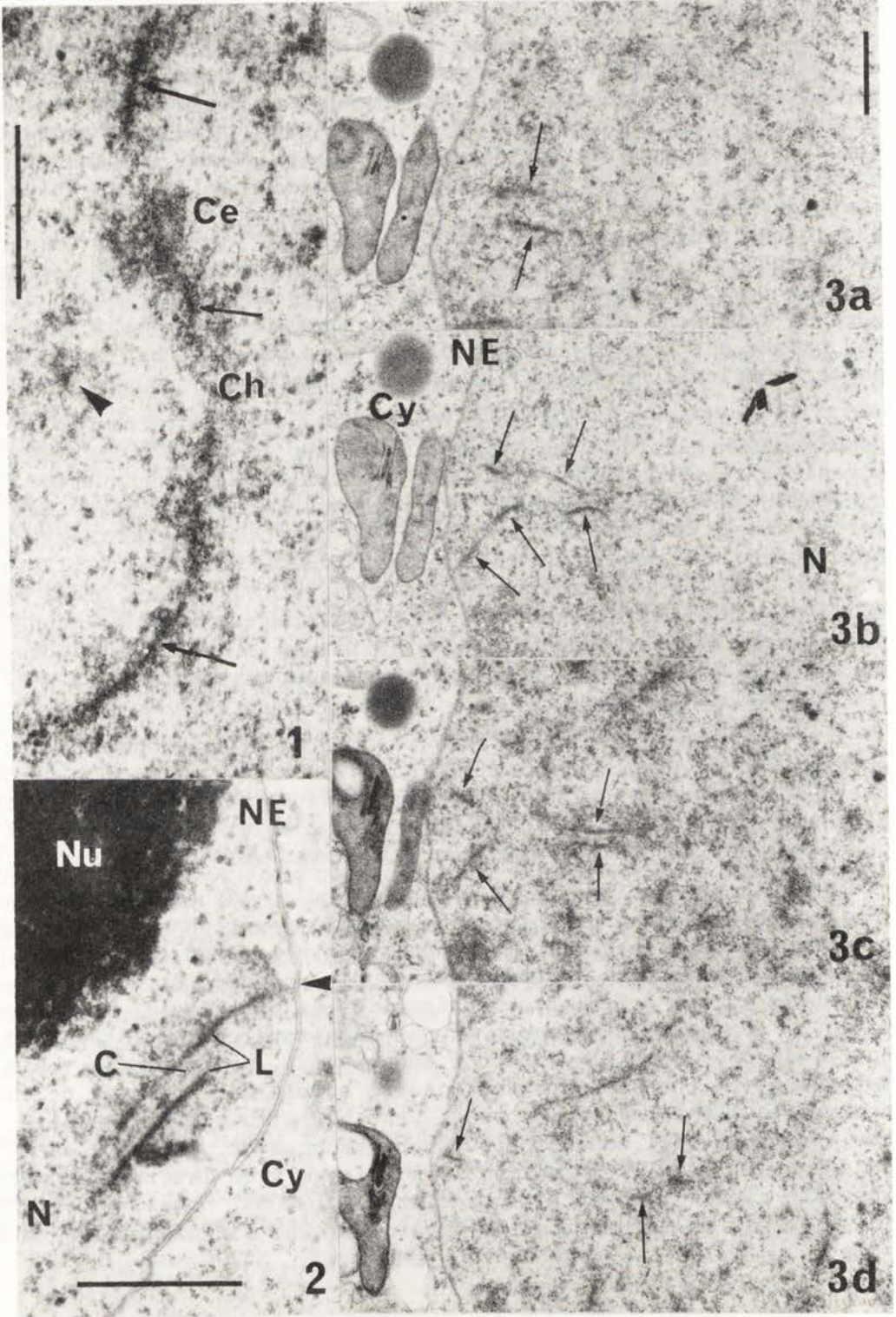


Figure 1. Unpaired leptotene chromosome with lateral component (arrows) in longitudinal section. Cross sectioned lateral component (large arrow head); Centromere (Ce); Chromatin (Ch). x 26,500 (Bar 1 μ).

Figure 2. Part of zygotene cell with synaptonemal complex near its end (arrow head) at the nuclear envelope (NE). Lateral components (L); Central component (C); Nucleoplasm (N); Cytoplasm (Cy); Nucleolus (Nu). x 25,000 (Bar 1 μ).

Figure 3a-d. Four serial sections of a partially paired zygotene bivalent. Lateral components (arrows) are unpaired at their attachment to the nuclear envelope (NE), but paired in the interior of the nucleus (c and d). Nucleoplasm (N); Cytoplasm (Cy). x 12,400 (Bar 1 μ).

the nuclear envelope means that synaptonemal complex formation often occurs first near the ends of chromosomes. Later, more initiations occur at other sites as matching improves. This is in agreement with the findings of Burnham et al. (1972), who studied pairing in maize translocation double heterozygotes. Knobs and centromeres do not appear to facilitate synaptonemal complex formation, but may aid in alignment. Lateral condensation of chromatin occurs but no shortening of lateral components is expected. The occurrence of matched ends in the haploid barley also suggests initiation occurs here, even though pairing is nonhomologous, and later in other parts of the chromosome, initiations seem to be at random (Figure 12). *Pachytene* is designated by the occurrence in the bivalents of fully formed synaptonemal complexes. Crossing over is believed to occur at this stage.

Synaptonemal complex and crossing over At pachytene in both *Neurospora* and maize it has been possible to trace the synaptonemal complex from one end to the other of each bivalent (Gillies, 1972, 1973a), and arrange them in pachytene karyotypes. In maize it was possible to identify individual bivalents (Gillies, 1973a). In normally paired bivalents in both cases there were no interruptions in the synaptonemal complexes, and they passed unaltered through the centromere, knobs and NORs.

In maize nuclei heterozygous for inversion 3b, it has been possible to reconstruct the lateral components and synaptonemal complexes of the inverted region of chromosome 3. It was found that in the presence of abnormal chromosome 10 (K10), a crossover enhancer (Rhoades, 1968), almost complete synaptonemal complex and loop formation occurs (Gillies et al., 1974). In the absence of K10 the pairing of lateral components to form synaptonemal complex, and consequently the inversion loop formation, was variable and much reduced. It seems inescapable that K10 acts to increase crossing over by improving the efficiency of pairing and synaptonemal complex formation. It is not, however, possible to

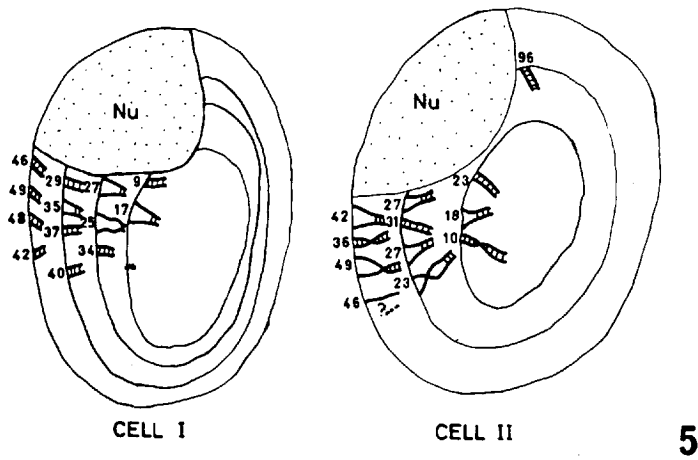
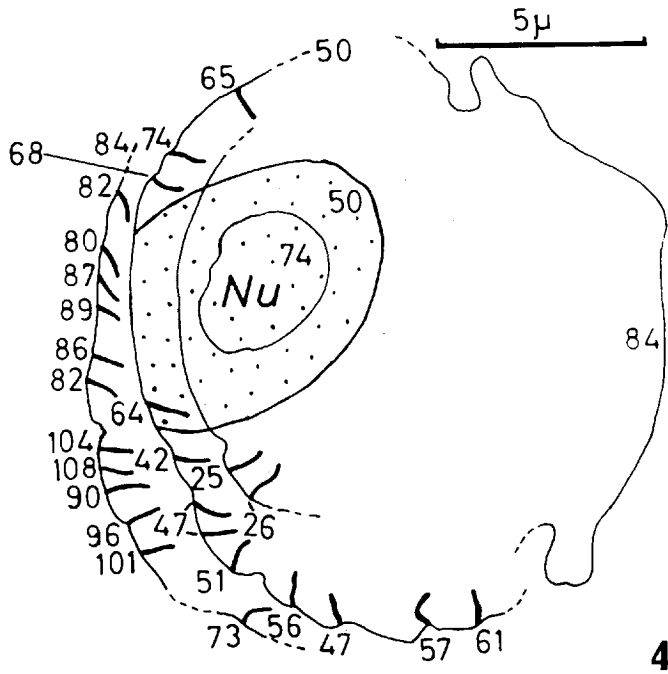
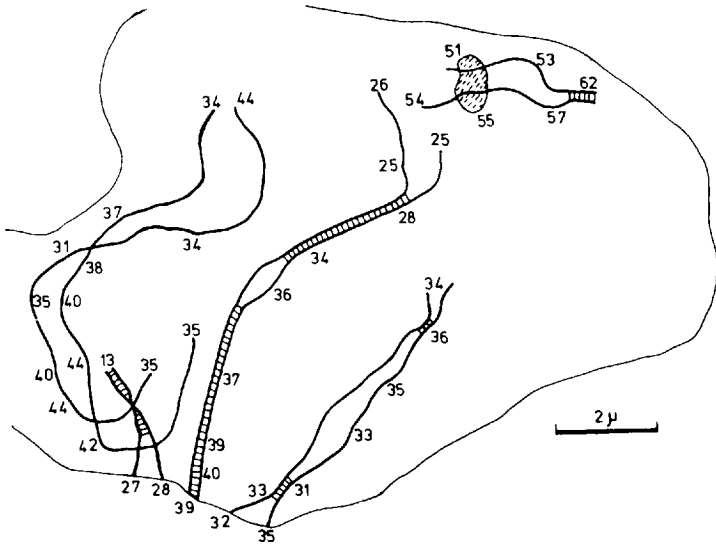
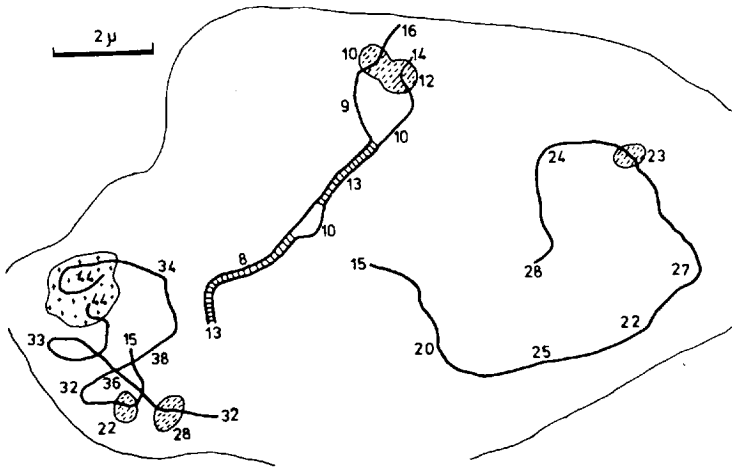


Figure 4. Reconstruction of leptotene nucleus showing points of attachment of lateral components (short heavy lines) to the nuclear envelope (thin lines) close to the nucleolus (Nu). Numbers are section numbers.

Figure 5. Diagrammatic representation of points of attachment of lateral components to the nuclear envelope in two zygotene nuclei. Synaptonemal complexes shown by cross striations. Numbers are section numbers. Nucleolus (Nu).

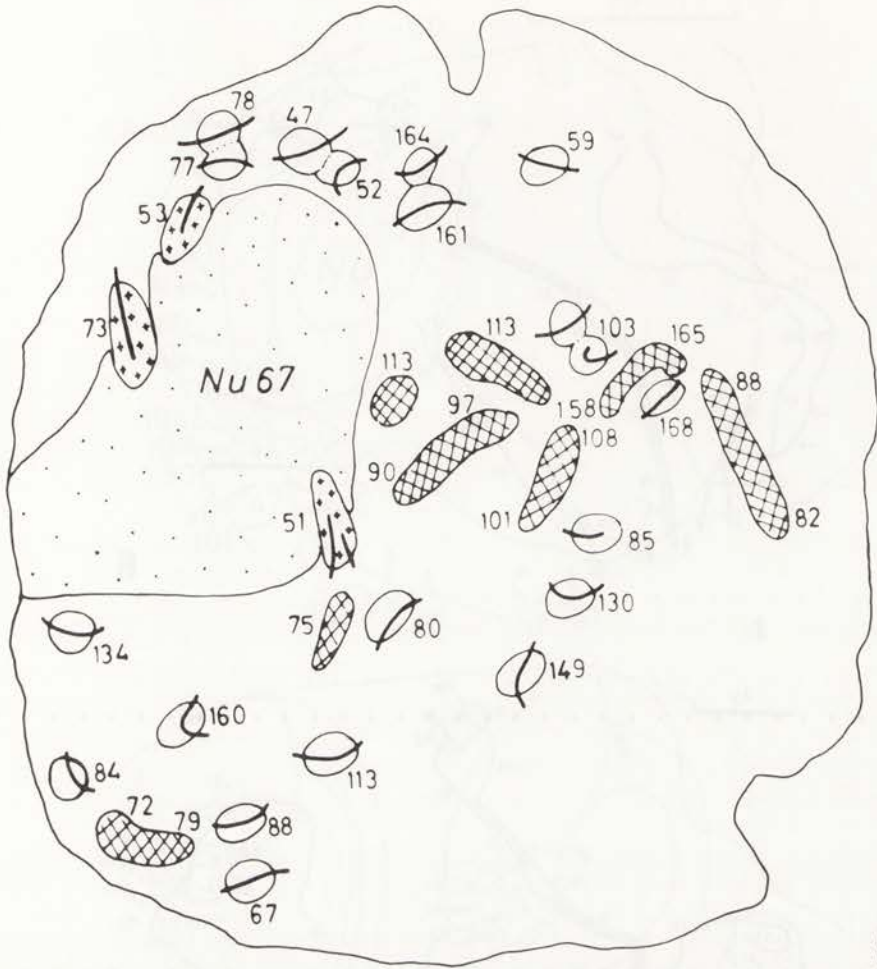


6

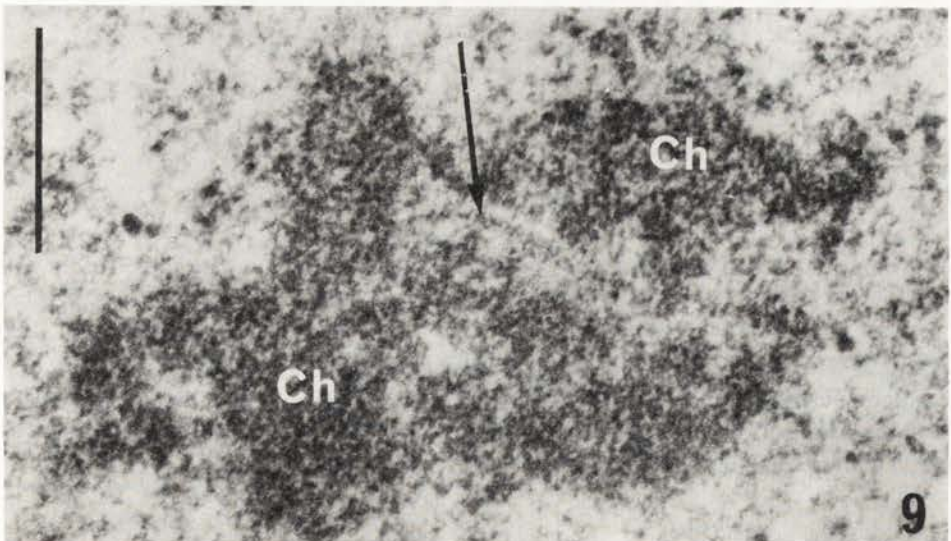


7

Figures 6 and 7. Reconstructions of portions of paired and unpaired lateral components from a zygotene nucleus. Synaptonemal complexes are cross striated. Centromeres are dashed. Nucleolus organizer region (NOR) is dotted. Numbers are section numbers. The incomplete outline of the nuclear envelope encloses the chromosomes.



8



9

Figure 8. Diagrammatic representation of positions of centromeres (open circles), chromatic knobs (cross hatched) and NORs (small crosses) in a leptotene nucleus. Lateral components represented by short heavy lines through the centromeres and NORs. Numbers are section numbers. Nucleolus (Nu).

Figure 9. Diplotene bivalent showing piece of central component (arrow) in space between two regions of chromatin (Ch). x 55,000 (Bar 0.5 μ).

distinguish between alignment effects acting *before* synaptonemal complex formation, and effects *during* the synaptonemal complex formation.

In the same lines of maize, apparently normal synaptonemal complexes have been found in the fold-back paired heteromorphic terminal segment of K10/N10 bivalents (Gillies et al., 1974). This pairing and synaptonemal complex formation is not homologous, since the fold-back occurs at different points in different bivalents.

Similar nonhomologous pairing and synaptonemal complex formation has been found in haploid barley. Up to 60% of the lateral components of the haploid complement (Figure 10) can be paired to form synaptonemal complexes (Figure 11), which seem identical with the synaptonemal complexes of diploid barley. Extensive stretches of synaptonemal complex have been formed, both between different chromosomes and by fold-back pairing in individual chromosomes (Figure 12). As mentioned earlier, there seem to be cases of chromosome ends being matched at the nuclear envelopes, but cases of unmatched ends not at the nuclear envelope also occurred (Figure 12). The pairing of other parts of the chromosomes appeared to be almost at random, and must be considered completely nonhomologous.

It is of interest to note that Sadasivaiah and Kasha (1971) have scored the chiasmata frequency in such haploid barley microsporocytes and found that by metaphase I there are no more than 0.05 chiasmata per nucleus. The implication is clearly that this extensive synaptonemal complex formation between nonhomologous chromosome parts does not lead to crossing over. It is reasonable to assume that after the gross alignment of chromosome ends at leptotene, and the fine alignment of homologues in synaptonemal complex formation, there is a final precise alignment of homologous DNA strands for the crossing over event in pachytene. This could be confined to selected regions (Stern and Hotta, 1973) and may be a mechanism for excluding crossing over from heterochromatic regions, centromeres etc.

K10 probably acts by ensuring that the second fine alignment step leads to synaptonemal complex formation between completely homologous regions. This would then increase the efficiency of the

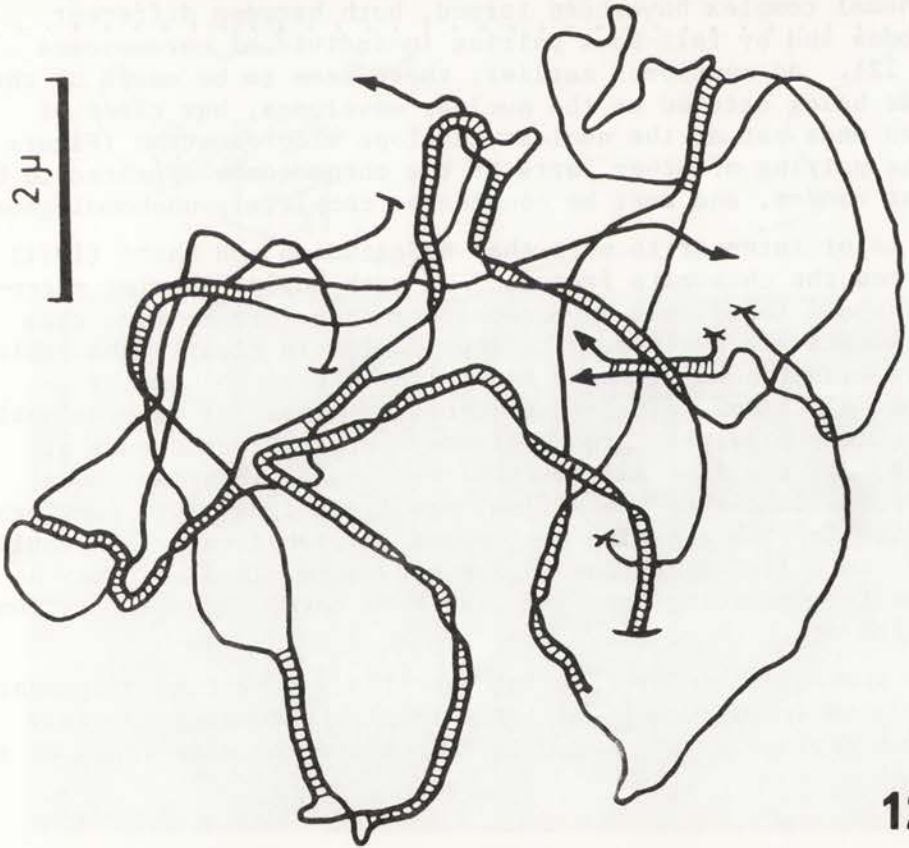
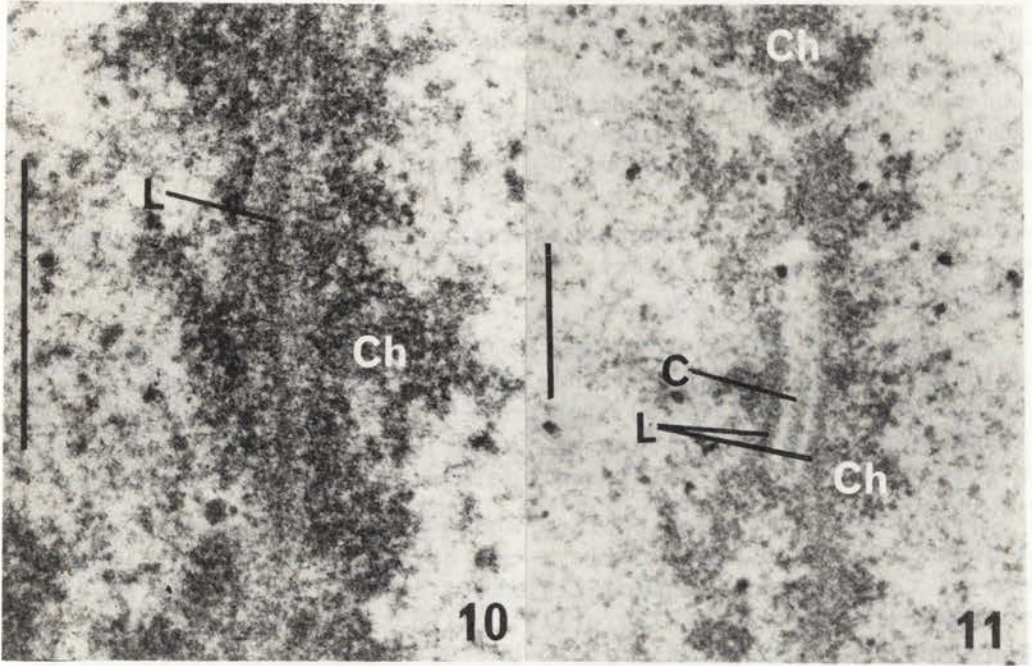


Figure 10. Unpaired lateral component (L) and associated chromatin (Ch) from pachytene stage in haploid barley. x 75,000 (Bar 0.5μ).

Figure 11. Synaptonemal complex from haploid barley at pachytene. Lateral components (L); Central component (C); Chromatin (Ch). x 40,000 (Bar 0.5μ).

Figure 12. Reconstruction of portions of lateral components and synaptonemal complexes (cross striated) from pachytene nucleus of haploid barley. Short perpendicular lines indicate chromosomes ending at nuclear envelope; arrow heads indicate lateral components continue off drawing; X indicates lateral component end lost.

third crossing over step. The precise DNA alignment cannot occur in nonhomologously paired chromosomes, and hence no crossing over occurs. However, the fact that extensive synaptonemal complex formation can occur in nonhomologous regions argues against synapsis being initiated by matching of DNA sites (Stern and Hotta, 1973).

Synaptonemal complex and chiasmata Examination of diplotene and diakinesis nuclei of maize revealed that the synaptonemal complexes disintegrate or are removed as the bivalents undergo rapid contraction. However, at early diplotene stages remnants of the synaptonemal complexes remain in some regions of bivalents (Figure 9). These regions usually retain a central component remnant in a central space of the approximate dimensions of the synaptonemal complex, but no sign of the lateral components is found. In a reconstructed diplotene nucleus in which the bivalents were about 40% of pachytene length, approximately 56 such remnants were found, but not all have well developed central component remnants remaining. The number per bivalent ranged from four to nine, with at least one in each bivalent arm. In diakinesis nuclei it was not possible to distinguish any synaptonemal complex remnants.

It appears that the synaptonemal complex disappears after only a small amount of diplotene chromosome contraction. Moens (1974) reports that in *Locusta* most synaptonemal complexes are absent after bivalents have contracted only about 20% from the full pachytene length. The delayed removal of some synaptonemal complex components in certain regions of diplotene bivalents has been reported in mouse (Solari, 1970), in *Neottiella* (Westergaard and von Wettstein, 1970b) and in mole cricket (Sotelo et al., 1973). In each case it has been suggested that the synaptonemal complex remnants are chiasmata precursors. The average frequency of synaptonemal complex components found here in maize at early diplotene (5.6 per bivalent) is about double the chiasmata frequency of 2.7 per bivalent reported by Darlington (1934) at a late diplotene stage. This suggests that half of the synaptonemal complex remnants could be potential chiasmata sites. As the chromosome contraction continues into diakinesis the chiasmata may develop fully into chromatin bridges. The synaptonemal complex remnants could then

be discarded as they became superfluous.

ACKNOWLEDGMENT

Work was carried out at the Institute of Genetics, University of Copenhagen, and the Carlsberg Laboratory Physiology Department, Copenhagen. It was assisted in part by the Danish Natural Sciences Research Council and the Carlsberg Foundation.

REFERENCES

- Burnham, C.R., J.T. Stout, W.H. Weinheimer, R.V. Kowles and R.L. Phillips. 1972. Chromosome pairing in maize. *Genetics* 71: 111.
- Darlington, C.D. 1934. The origin and behaviour of chiasmata. VII. *Zea mays*. *Z. indukt. Abstamm. Vererbung.* 67: 96.
- Gillies, C.B. 1972. Reconstruction of the *Neurospora crassa* pachytene karyotype from serial sections of synaptonemal complexes. *Chromosoma* 36: 119.
- . 1973a. Ultrastructural analysis of maize pachytene karyotypes by three dimensional reconstruction of the synaptonemal complexes. *Chromosoma* 43: 145.
- . 1973b. Methods for electron microscopic study of maize pachytene chromosomes. *Maize Genetics Coop. Newsletter* 47: 25.
- . 1974. The nature and extent of synaptonemal complex formation in haploid barley. *Chromosoma*. In press.
- Gillies, C.B., S.W. Rasmussen and D. von Wettstein. 1974. The synaptonemal complex in homologous and non-homologous pairing of chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38: 117.
- Lu, B.C. 1970. Genetic recombination in *Coprinus*. II. Its relations to the synaptonemal complexes. *J. Cell. Sci.* 6: 669.
- Moens, P.B. 1969. The fine structure of meiotic chromosome polarization and pairing in *Locusta migratoria* spermatocytes. *Chromosoma* 28: 1.
- . 1974. Quantitative electron microscopy of chromosome organization at meiotic prophase. *Cold Spring Harbor Symp. Quant. Biol.* 38: 99.
- Rhoades, M.M. 1968. Studies on the cytological basis of crossing over. In Replication and recombination of genetic material. (W.J. Peacock and R.D. Brock, ed.) p.229. Aust. Acad. Sci., Canberra.
- Sadasivaiah, R.S. and K.J. Kasha. 1971. Meiosis in haploid barley - an interpretation of non-homologous chromosome associations. *Chromosoma* 35: 247.
- Solari, A.J. 1970. The behaviour of chromosomal axes during diplotene in mouse spermatocytes. *Chromosoma* 31: 217.

- Sotelo, J.R., R.B. Garcia and R. Wettstein. 1973. Serial sectionin study of some meiotic stages in *Scaptericus borrelli* (Grylloidea). *Chromosoma* 42: 307.
- Stern, H. and Y. Hotta. 1973. Biochemical controls of meiosis. *Ann. Rev. Genet.* 7: 37.
- Westergaard, M. and D. von Wettstein. 1970a. The nucleolar cycle in an Ascomycete. *C.R. Lab. Carlsberg* 37: 195.
- _____, _____. 1970b. Studies on the mechanism of crossing over. IV. The molecular organization of the synaptinemal complex in *Neottiella* (Cooke) Saccardo (*Ascomycetes*). *C.R. Lab. Carlsberg* 37: 239.
- _____, _____. 1972. The synaptinemal complex. *Ann. Rev. Genet.* 6: 71.

Replication initiation in yeast chromosomes

Wylie Burke, Carol Shaw Newlon
and Walton L. Fangman

Induction of mutation by N-methyl-N'-nitro-N-nitrosoguanidine (NG) appears to be replication dependent in yeast since mutation rates were lower in G1 cells than in S period cells or in cells of an asynchronous population. In cells going through a synchronized S period markers exhibited discrete and distinct times of maximal mutation by NG. This result suggests that DNA sequences within the small yeast genome (9×10^9 daltons) replicate in a defined temporal order. Initiation of DNA replication must therefore involve specific regions where replication occurs at predetermined times during the S period.

The enormous size of eukaryotic chromosomal DNA molecules poses a serious obstacle to the study of their structure. There are, however, eukaryotic organisms with very small chromosomes. The yeast *Saccharomyces cerevisiae*, with a genome of approximately 9×10^9 daltons of DNA distributed among 17 chromosomes, is one of these. The size of chromosomal DNA released from yeast spheroplasts by lysis has been measured by both sedimentation velocity (Petes and Fangman, 1972; Blamire et al., 1972) and electron microscopy (Petes et al., 1973). The results suggest that the average yeast chromosome contains a single DNA duplex with a mass of about 5×10^8 daltons, and that individual chromosomal DNA molecules range in size from about 10^8 to 10^9 daltons. In spite of their small size, yeast chromosomes replicate in a manner much like that found in higher eukaryotic organisms (Newlon et al., 1974). Initiation of replication occurs at multiple sites along a DNA molecule, and initiation sites are closely spaced at about 25 micron (0.5×10^8 dalton) intervals. During an S period, initiation occurs at approximately 200 sites.

The precision with which initiation sites for chromosome replication are chosen in each cell cycle in eukaryotic organisms is not known. Studies on the temporal sequence of DNA replication during the S period indicate that initiation sites in eukaryotes cannot be chosen completely at random (e.g., Muldoon et al., 1971; Stubblefield, 1966). We are attempting to determine whether the

same sites are used for the initiation of replication during each yeast cell cycle, and if so whether a given site is always activated at the same time within the S period. Our approach involves an analysis of the precision of replication time for individual genes. Results of this approach allow deductions about the specificity of replication start sites. An analysis by electron microscopy of small DNA molecules isolated as replication intermediates has also been started.

METHODS

Strains and growth conditions For sedimentation experiments a strain (D5 ρ -6) which lacks mitochondrial DNA was used. It was derived from a prototrophic diploid, A364A D5 U⁺A⁺ (Newlon et al., 1974) by the procedure of Goldring et al. (1970). The cells were grown in Y-minimal medium (6.7 g Difco yeast nitrogen base without amino acids, 10 g succinic acid, and 6 g NaOH per liter) supplemented with 20 g glucose and 0.5 g Difco yeast extract per liter. Uracil-6-³H (1 mCi/.042 mg, New England Nuclear) was added to give 10 μ Ci/ml. Since the yeast extract (0.05%) supplies about 1 μ g/ml uracil, the uracil specific activity was about 10 μ Ci/ μ g. The electron microscopy experiments employed a haploid strain, A364A, which has the mating type α and requires uracil. Growth and labeling were as above except that an additional 2 μ g/ml uracil was added to the medium.

For the mutagenesis experiments, the strains were 1422-11D (α , *leu-1*, *ade-1*, *his-2*, *met-2*, *lys-1*, *trp-5*, *ura-3*, *gal-1*) and 3163 (α , *leu-1*, *met-1*, *ade-5*, *ura-3*, *trp-3*) an Arg⁺ derivative of X3163-4C, both supplied by D. Hawthorne. With the exception of *his-2*, the markers in 1422-11D are not suppressed by known super-suppressors (D. Hawthorne, personal communication). Cells were grown in YM-1 medium (5 g Difco yeast extract, 10 g bacto-peptone, 6.7 g yeast nitrogen base with amino acids, 10 g succinic acid, 6 g NaOH per liter) supplemented with 50 mg adenine, 50 mg uracil and 20 g glucose per liter. For radioactive labeling, uracil-6-³H was added to give 10 μ Ci/ml in medium without the uracil supplement. The yeast extract (0.5%) in the medium provides about 10 μ g/ml uracil.

Spheroplast formation and DNA sedimentation Yeast cells (10^7 to 10^8) were filtered (Gelman GA6), washed with cold H₂O and suspended in 4 ml 1.2 M sorbitol-0.2 M Tris-0.02 M Na EDTA-0.1 M 2-mercapto-ethanol (pH unadjusted). After incubation at room temperature for 10 min, the cells were filtered, washed with 1 M sorbitol-0.1 M sodium citrate-0.01 M Na EDTA, pH 5.8 (SCE) and suspended in 0.2 ml SCE. Ten μ l Glusulase (Endo Laboratories) was added and the mixture incubated at 37° for 15 min. Two hundred μ l of the spheroplast suspension were layered onto a 12 ml linear 15-30% sucrose gradient

containing 0.01 M Na EDTA (pH 8), 0.01 M Tris HCl (pH 8) and 1 M NaCl, previously layered with 100 μ l T4 14 C-DNA (Petes and Fangman, 1972) in 5% sodium lauroyl sarcosinate (sarkosyl) in SCE. After layering the spheroplasts an additional 100 μ l 5% sarkosyl in SCE was layered onto the gradient. The gradient tubes were allowed to stand 15 min at room temperature (21°), then one hour at 5° before centrifugation at 5° in a Spinco SW41 rotor. For electron microscopy 10^8 spheroplasts were layered on the gradients. For all other sedimentation experiments 10^7 spheroplasts were layered. Gradient samples were treated with NaOH to hydrolyze RNA (Petes and Fangman, 1972).

Electron microscopy After centrifugation, gradients were collected (Petes et al., 1973) in about eight fractions. Appropriate fractions, determined by parallel sedimentation of T4 DNA, were concentrated in dialysis tubing surrounded by Bio-Gel P-10 (BioRad Laboratories) and prepared for electron microscopy as described previously (Petes et al., 1973).

α -factor arrest and synchronous cultures α -factor was concentrated from α strain supernatants by the column procedure of Duntze et al. (1973), except that before final elution of the factor with 0.01 M HCl - 80% ethanol, the column was washed with 0.001 M HCl - 80% ethanol. The column eluant was concentrated by evaporation in the cold. α -factor was assayed by monitoring the accumulation of unbudded cells in an asynchronous population to which it was added. For α -factor arrest and synchrony experiments, α -factor was used at a concentration which maintained 95% unbudded cells for at least 30 min longer than a generation time.

For α -factor arrest populations, α -factor was added to an asynchronous culture at 2.5×10^7 cells/ml. Cells were collected after a time period equivalent to one generation of growth. For synchronous cultures, arrested cells were collected by filtration, washed with fresh medium and resuspended in fresh medium. Cells were sonicated after α -factor treatment to break up cell clumps. When incorporation of radioactive uracil was determined in cultures, samples were taken into cold TCA (10% TCA final) containing 1 mg/ml uracil, centrifuged in the cold and resuspended in 0.6 N NaOH containing 1 mg/ml uracil. Samples were incubated at 37° overnight, then chilled and precipitated with an equal volume of 20% TCA, 0.6 N HCl. Filtering and counting of radioactive samples were as previously described (Petes and Fangman, 1972).

Mutagenesis Cell samples taken for N-methyl-N¹-nitro-N-nitrosoguanidine (NG) mutagenesis were chilled and then collected by centrifugation or filtration. The samples were washed with cold sterile distilled water, and then resuspended in 0.2 M sodium acetate,

pH 5.0, containing 1 mg/ml NG (K & K Co.). Cells at 5×10^8 /ml were either aerated or shaken in a gyratory shaker during mutagenesis. After incubation, cells were collected by filtration, washed with cold sterile distilled water, and then resuspended in sterile distilled water for plating. Cells were spread on agar (2%) plates made of Y-minimal medium supplemented with 20 g glucose and 50 mg each of the appropriate nutrients per liter. For experiments with strain 1422-11D, mutagen treatment was carried out at 0°. For strain 3163, it was necessary to carry out the mutagen treatment at 30° in order to obtain a measurable rate of reversion to Ade⁺.

Genetic analysis of revertants was carried out by standard techniques (Mortimer and Hawthorne, 1969).

RESULTS

Replication dependent mutagenesis If DNA sequences are associated with ordered initiation sites then replication of these sequences will be ordered. We have thus examined whether replication of different genes is ordered during the S period. The time of replication of a gene was determined by a procedure developed with bacteria. It has been shown in *Escherichia coli* that the frequency of mutation induced by N-methyl-N'-nitro-N-nitrosoguanidine (NG) is maximal for a genetic marker at or near the time the marker is replicated. In bacteria the time of maximal mutation induction for different genes correlates with their time of replication determined by independent physical techniques (see Hohlfield and Vielmetter, 1973). If NG-induced mutagenesis is correlated with replication in yeast then the time of maximal induction of mutation by NG can be used to deduce the time of replication of individual genes.

As in bacteria, NG appears to act preferentially on yeast DNA during replication. This can be shown by comparing the rates of NG mutagenesis with growing cells and cells arrested in the G1 phase of the cell cycle by α -factor (Bücking-Throm et al., 1973). Table 1 presents the rates of reversion for three auxotrophic markers; all of them showed a pronounced reduction in the NG-induced rates for G1 arrested cells compared to asynchronously growing cells. For two markers, *leu-1* and *met-2*, the NG-induced rates of reversion in G1 arrested cells were only slightly above the spontaneous frequencies. For the third marker, *ade-1*, the rate of reversion in G1 arrested cells was well above the spontaneous frequency but in several experiments was always less (4- to 8-fold) than the rate observed in growing cells. We conclude that in yeast, as in bacteria, NG probably preferentially mutagenizes replicating DNA.

This conclusion is also supported by results with cycloheximide-treated cells. Cycloheximide, a protein synthesis inhibitor,

has been shown to inhibit the initiation of the S period (Hereford and Hartwell, 1973; Williamson, 1973). Cells which have started DNA synthesis, however, complete the S period in the presence of the inhibitor. We have observed a reduction (5- to 15-fold) in NG-induced rates of reversion in cells previously treated with 200 $\mu\text{g/ml}$ cycloheximide for a time (40 min or longer) sufficient to allow completion of the S period.

TABLE 1
Induction of revertants

Marker	Spontaneous frequency*	Nitrosoguanidine-induced rate [†] (revertants per 10 ⁸ viable cells)	
		Asynchronous culture	α -factor arrested (G1) cells
<i>ade-1</i>	10	2,300	600
<i>leu-1</i>	1	270	4
<i>met-2</i>	5	140	14

*Values were similar for asynchronous and α -factor arrested cells.

†Treatment was at 0° for 30 min. There was less than 20% lethality.

Mutagenesis of synchronous cultures The following experiments demonstrate that maximal NG mutagenesis occurs during the S period, and test the idea that the replication of individual genes is ordered during the DNA synthetic period. In these experiments, mutagenesis was performed with cells undergoing a synchronous S period. Yeast cells of *a* mating type were arrested in the G1 period with α -factor, collected by filtration and suspended in fresh medium without α -factor. Cells treated in this way undergo a partially synchronous DNA synthesis period as shown by the incorporation of ³H-uracil into DNA (Figure 1). NG mutagenesis treatments were carried out with samples taken from such a synchronous population of cells. Figure 2 shows the results expected for this experiment based on two general models for the timing of replication of a gene. In one model the time of replication of the gene is *random*. Different cells in the population replicate the gene in question at different times. Hence the time of maximal mutation by NG for the total population corresponds to the entire S period. In the second model the time of replication of the gene is *precise*, so that each cell replicates the gene at the same relative time within the S period. In this case

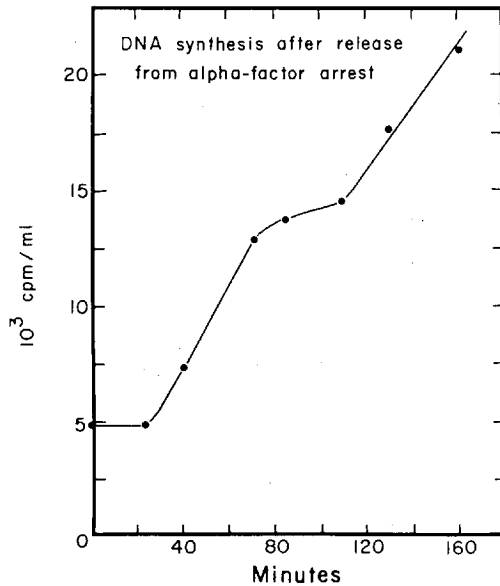


Figure 1. Incorporation of ^3H -uracil into DNA in a culture released from α -factor arrest. Cells of strain 1422-11D were pre-labeled with ^3H -uracil for about 0.3 generations prior to α -factor arrest and treated with α -factor in the continued presence of label. After release, additional label was added at the same specific activity. Samples were taken at 20 min intervals for determination of alkali-stable incorporation.

the time of maximal mutation by NG is discrete, occurring over a short interval, the length of which is determined by the degree of asynchrony of the population. Figure 3 shows the rates of reversion to Ade^+ , Leu^+ , and Met^+ during the course of a partially synchronous S period. All three show distinct peaks of NG-induced reversion during synchronous growth. Moreover, the peaks are separable, Ade^+ revertants occurring maximally about 10 min before Leu^+ , and Met^+ revertants occurring maximally about 10 min after Leu^+ . The mutation frequency declines at the end of the first S period and then rises again as the culture enters a second, less synchronous, S period. These results indicate that NG mutagenesis is replication dependent, and the peaks of reversion are presumed to represent the times during the S period when the individual markers are being replicated. The timing of appearance of the three peaks provides evidence that gene replication in yeast is ordered. The three genes, *ade-1*, *leu-1* and *met-2*, are unlinked; their chromosomal locations are shown in Figure 4.

Genetic testing was carried out to determine whether the NG-induced revertants were the result of back mutation at the original

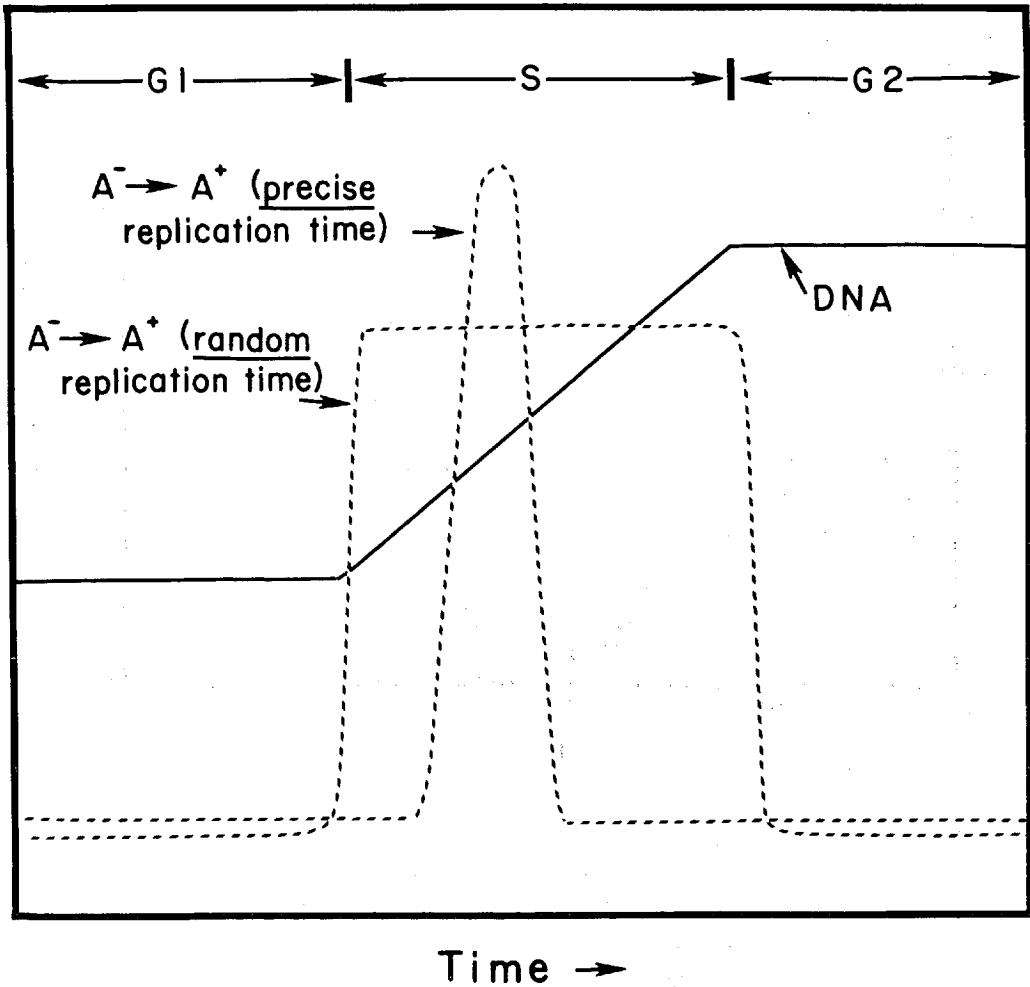


Figure 2. Data predicted for two models for the timing of replication of a gene. Cells are removed at frequent intervals from a culture undergoing a synchronous S period, treated with NG and plated to determine the frequency of A^+ revertants.

site or were caused by unlinked suppressors. In the former case, the peaks observed in Figure 3 would represent replication times for the genes *ade-1*, *leu-1* and *met-2*; in the latter case, the peaks would represent replication times for suppressor loci. For each marker, twenty revertants were taken from a peak sample. Each revertant was crossed to an α strain which was wild type for the marker in question. The diploid was sporulated and a random spore analysis performed on the spores. If an NG-induced revertant is a result of an unlinked suppressor, 25% of the spore products will be

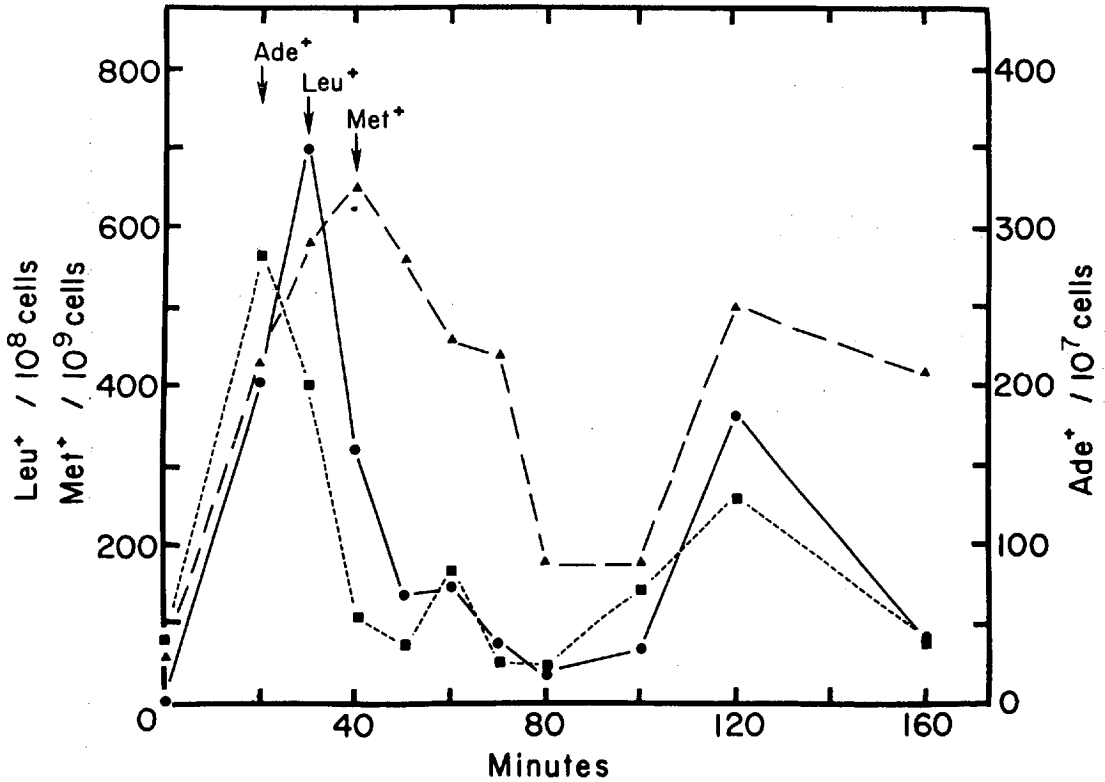


Figure 3. NG-induced rates of mutation to Ade⁺, Leu⁺ and Met⁺ during a synchronous S period. Samples were taken from a synchronous culture of 1422-11D at 10 min intervals, treated with NG (30 min at 0°), and plated to determine the frequency of Ade⁺, Leu⁺ and Met⁺ colonies.

auxotrophic. In such tests Leu⁺ and Ade⁺ revertants generated no auxotrophic spores among 60-90 spores examined for each revertant. Therefore, the NG-induced mutations must be at, or very closely linked to, the original mutant sites.

The Met⁺ revertants fell into two classes. Half of them produced no Met⁻ spores. The other half yielded 25% Met⁻ spores, indicating that in these revertants the Met⁺ phenotype is the result of a suppressor unlinked to *met-2*. Therefore, the peak of NG-induced Met⁺ revertants results from mutations at the *met-2* locus and one or more suppressor loci. This result probably accounts for the fact that the Met⁺ peak is broader than either the Leu⁺ or Ade⁺ peak.

Another yeast strain carried two NG-revertible markers on the same chromosome: *leu-1* and *ade-5*. Both markers are on the left arm of chromosome VII, *leu-1* located close to the centromere and *ade-5* located more distally (Figure 4). Results of a synchronous culture

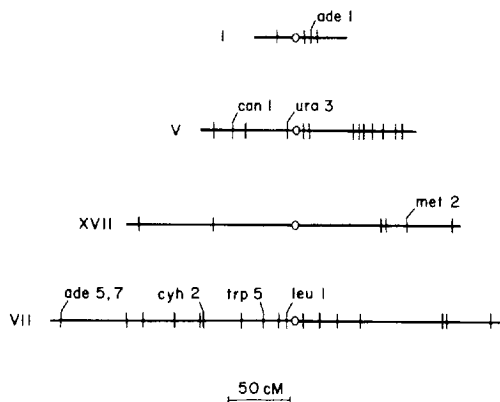


Figure 4. Genetic maps of four yeast chromosomes. The vertical lines represent positions of mapped genes; a few of the genes are designated by name. The scale is in centimorgans. For chromosome XVII the distance from the centromere to the first marker on each arm is unknown. Most of the mapping positions are taken from the data of Mortimer and Hawthorne (1973).

experiment with this strain are shown in Figure 5. Peaks of reversion to Leu^+ and Ade^+ occur about 20 min apart, suggesting that replication of the left arm of chromosome VII takes at least 20 min to complete. A third marker, *can-1*, was studied in this strain by measuring the induction of forward mutations to canavanine-resistance. This marker shows a peak of mutation at about the same time that the peak of Leu^+ revertants occurs. Genetic testing of these mutants to establish the sites of mutation is in progress.

Since the *leu-1* marker is present in both of the strains used in these experiments, the apparent order of replication of the five markers studied can be summarized by placing them on a scale where the time of replication of *leu-1* is given the normalized time of 0 min (Figure 6).

Isolation of small chromosomal DNA molecules The location of replication start sites could be determined if a series of replication intermediates for a specific chromosomal DNA was examined in the electron microscope. To obtain a collection of intermediates we are isolating the smallest yeast chromosomal DNA molecules. Specific molecules are partially resolved by sedimentation (Blamire et al., 1972; Finkelstein et al., 1972; Petes et al., 1974). This allows recovery of DNA molecules which fall within a restricted range of sizes (Petes et al., 1973), which must correspond to a small number of chromosomes. We are concentrating on the smallest molecules because they are easiest to analyze in the electron microscope.

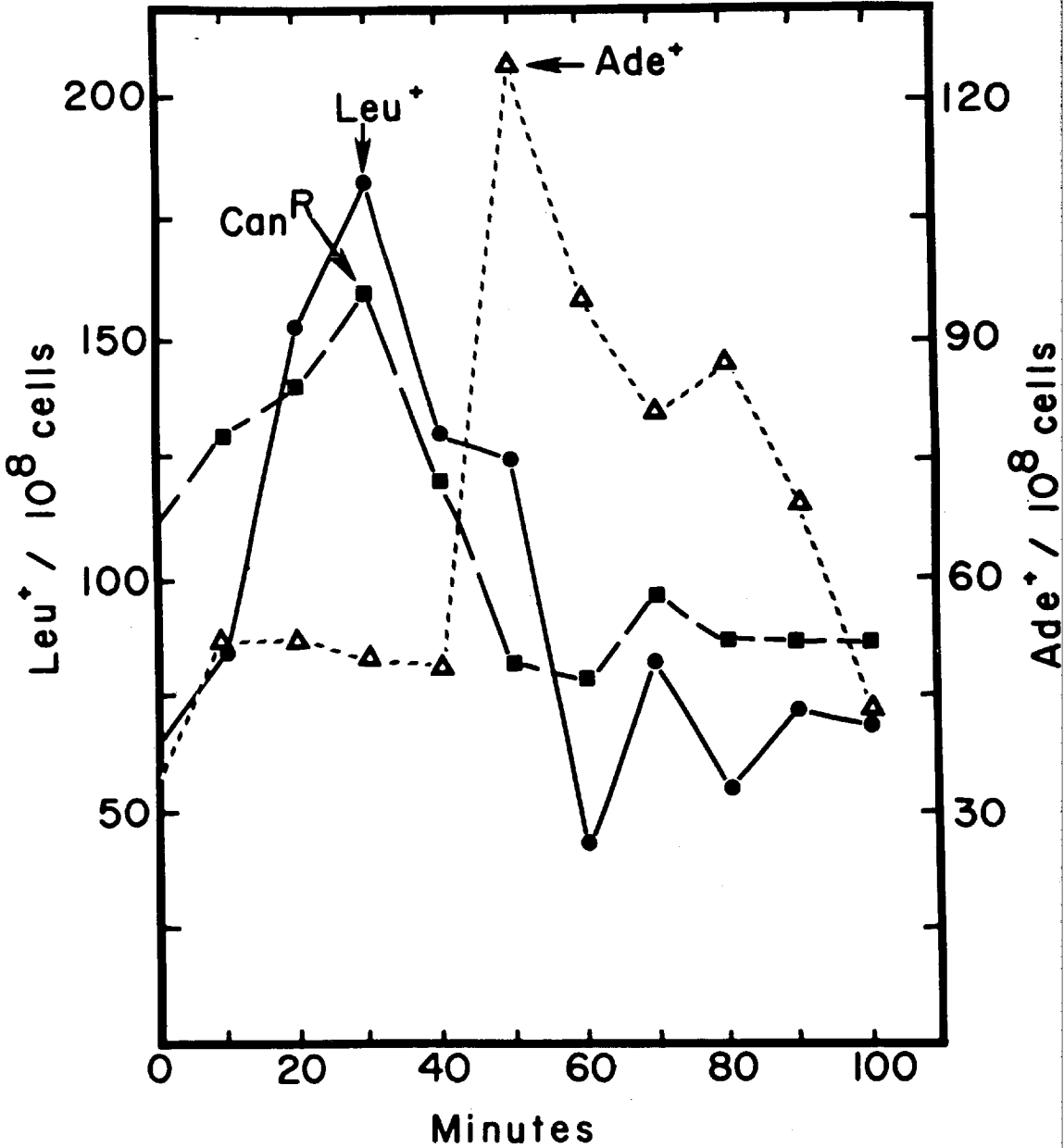


Figure 5. NG-induced rates of mutation to Can^R, Leu⁺ and Ade⁺ during a synchronous S period. Samples were taken from a synchronous culture of 3163 at 10 min intervals, treated with NG (30 min at 30°), and then plated to determine the frequency of Leu⁺, Ade⁺ and Can^R colonies. Canavanine was employed at 60 μg/ml.

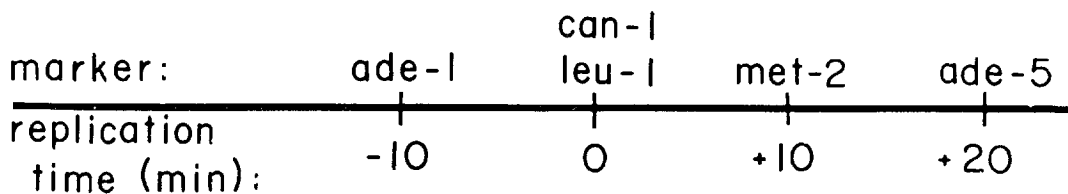


Figure 6. Time of maximal mutagenesis for five markers. The time of maximal mutation for a marker is assumed to represent the time at which a particular marker is replicated. Mutation to Leu^+ is assumed to occur at the same time in the two strains employed, and is arbitrarily taken as 0 min.

The sedimentation coefficient of high molecular weight yeast chromosomal DNA is a function of rotor speed (Petes and Fangman, 1972). An inverse relationship between rotor speed and sedimentation velocity for DNA molecules of molecular weight greater than 10^8 has been observed in several laboratories (see Chia and Schumaker, 1974). This effect is demonstrated in Figure 7, which shows the patterns obtained when T4 DNA and yeast DNA were sedimented at two rotor speeds, for lengths of time such that $rpm^2 \times time$ was constant. On centrifugation at 10,000 rpm for 27 hours the bulk of the yeast DNA exhibited a sedimentation velocity which was greater than that of T4 DNA (Figure 7, Panel A), whereas when centrifugation was carried out at 30,000 rpm much of the yeast DNA sedimented more slowly than the T4 DNA marker (Figure 7, Panel B). The lower sedimentation velocity at 30,000 rpm is not the result of fragmentation or some other permanent change in the yeast DNA. When a gradient which had been centrifuged at 30,000 rpm was subsequently centrifuged at 10,000 rpm the yeast DNA increased in sedimentation velocity and, in fact, passed the T4 DNA (Figure 7, Panel D). Also the rotor speed effect cannot be explained as the result of a change in the properties of either the gradient or the DNA during the longer centrifugation at 10,000 rpm. When material which had been centrifuged at 10,000 rpm was subsequently centrifuged at 30,000 rpm the yeast DNA decreased in sedimentation velocity and the T4 DNA 'caught up' (Figure 7, Panel C). This conclusion is also supported by experiments in which both gradients and lysed material were allowed to stand in the cold for 27 hours before centrifugation.

Somewhat better resolution of small and large DNA molecules can be obtained with low speed (Figure 8) compared to high speed centrifugation (Figure 9). However, high speed centrifugation is advantageous because it allows DNA molecules to be prepared for electron

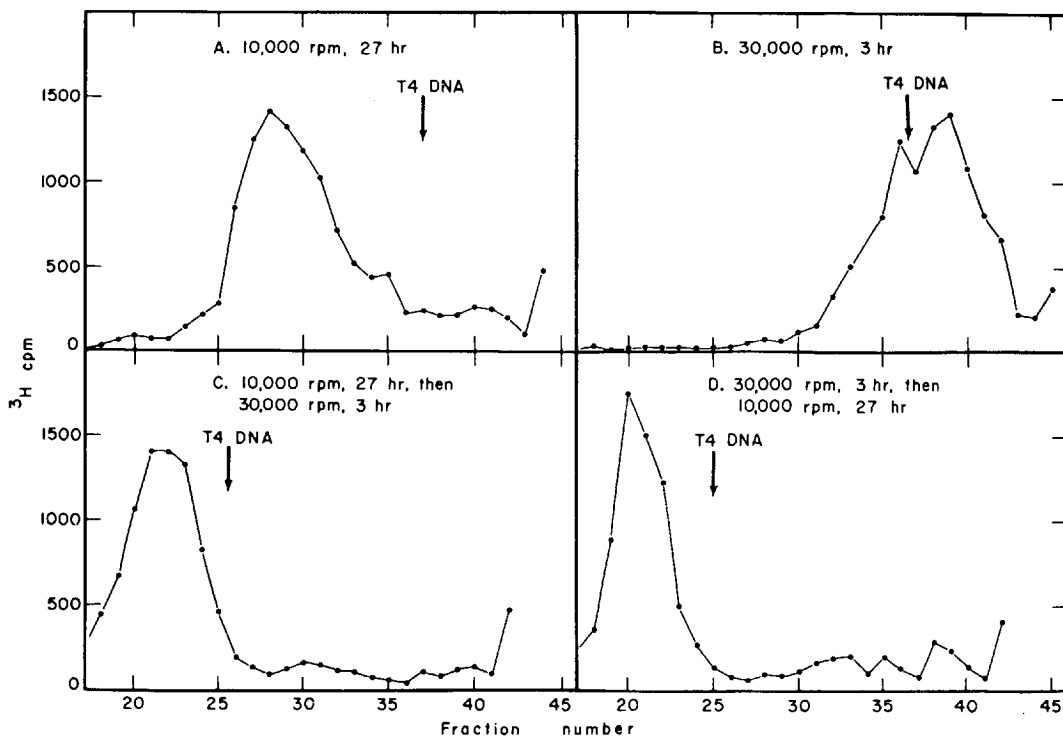


Figure 7. Effect of rotor speed on the sedimentation velocity of yeast DNA. For the gradients shown in panels C and D, the tubes were centrifuged at the first rotor speed shown. The rotor speed was then changed for further centrifugation. Under the conditions of centrifugation the sucrose gradient is isokinetic for 80% of the distance from top to bottom of the tube. The first 16 fractions contained little DNA and are not shown.

microscopic studies within a relatively brief time after cell lysis.

Replication structures in small molecules We are beginning a study of the replication of chromosomal DNA molecules in the size range of approximately 1 to 2×10^8 daltons. Yeast cells of α mating type are synchronized by incubating them with α -factor and subsequently removing it by filtration. When incubated in the presence of 0.1 M hydroxyurea to inhibit the rate of DNA chain propagation (Slater, 1973) replication intermediate molecules accumulate which contain small replication structures. As with larger DNA molecules (Newlon et al., 1974) the T4 size DNA molecules that we observe contain more than one replicating structure and hence more than one replication start site. An example is shown in Figure 10 which is a tracing of an electron micrograph of a molecule 50 microns in length which contains two replicating structures. Comparison of a series of

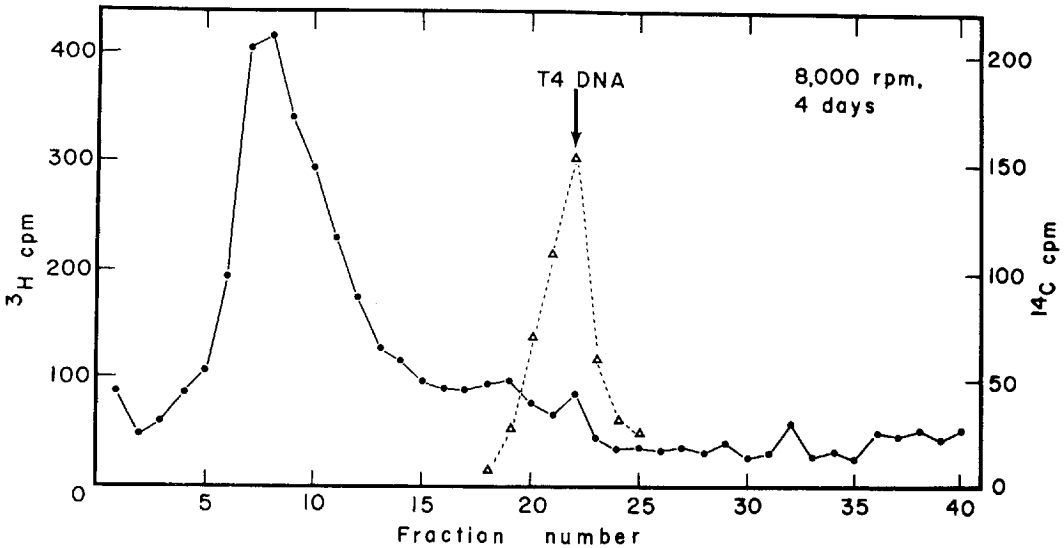


Figure 8. Sedimentation of yeast DNA at low rotor speed (8,000 rpm) for 4 days

replication intermediates for a particular molecule should allow us to reach conclusions about the specificity of replication initiation sites (Schnös and Inman, 1970).

DISCUSSION

In yeast cells going through a synchronous S period different genetic markers show distinct peaks of NG-induced mutation at different times. We take this as evidence that a particular DNA sequence replicates during a limited and programmed time in the S period. Density labeling techniques have previously been used to demonstrate that a temporal order of replication exists in the S period in animal cells (Mueller and Kajiwara, 1966) and in the simple eukaryote *Physarum* (Braun and Wili, 1969). With *Physarum*, it was possible to demonstrate that DNA sequences which replicated during one tenth of the S period in one cell cycle, replicated during the same fraction of a subsequent S period (Muldoon et al., 1971). The *Physarum* haploid genome consists of about 1.8×10^{11} daltons of DNA (Mohberg and Rusch, 1971), or 20-fold more DNA than the yeast genome. Therefore, the amount of DNA which has been shown to replicate in a precise temporal order in *Physarum* is equivalent to two yeast genomes. Our data thus extend previous observations by demonstrating a temporal order for the replication of DNA sequences within a very small eukaryotic genome.

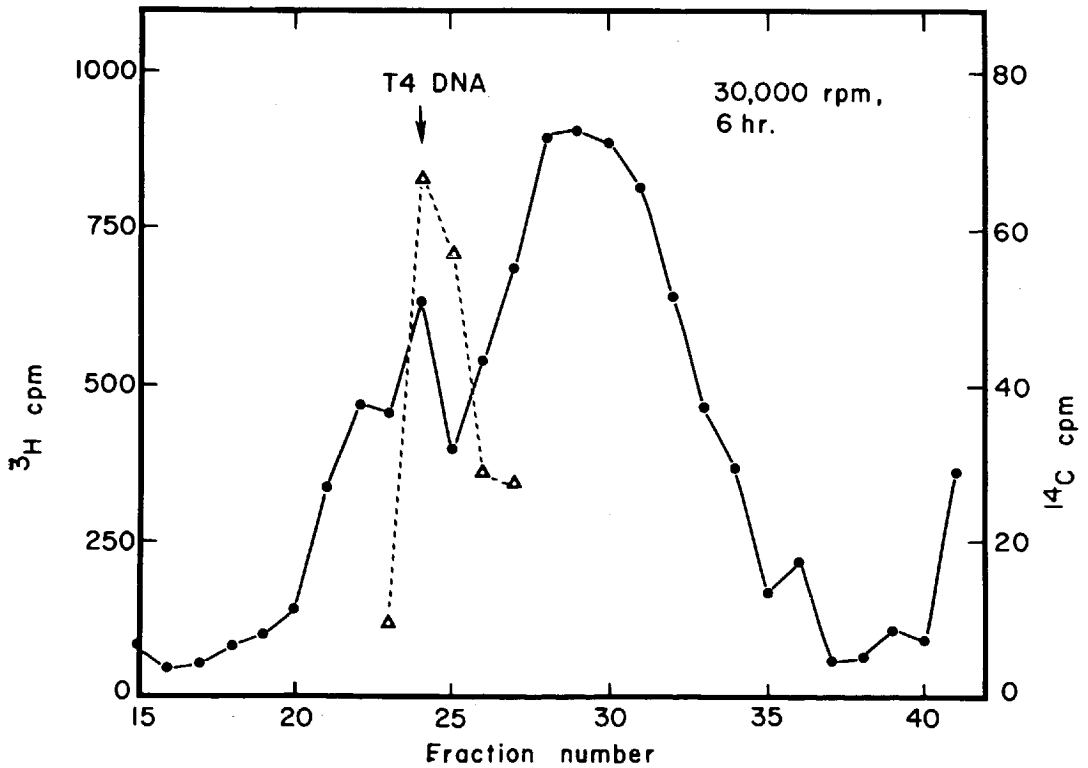


Figure 9. Sedimentation of yeast DNA at high rotor speed (30,000 rpm) for 6 hours

The observation that individual genes replicate at discrete times during the S period leads to the conclusion that initiation of replication is a fairly specific process, involving defined start regions that are activated at predetermined times in the S period. In the synchronous yeast cultures mutation of an individual gene appears to occur over a period of 20 min or more. A large proportion of the spread in the time of mutation most likely results from imperfect synchrony in the α -factor treated cultures. Initiation of new buds occurs over a period of time similar in duration to the spread observed for reversion of the *ade-1* and *leu-1* genes. Our data are, however, consistent with a limited amount of variation in both timing and location of initiation.

It should be possible to determine the degree of precision of initiation by a direct examination of replication intermediates. Techniques applied to virus DNA molecules (Schnös and Inman, 1970)

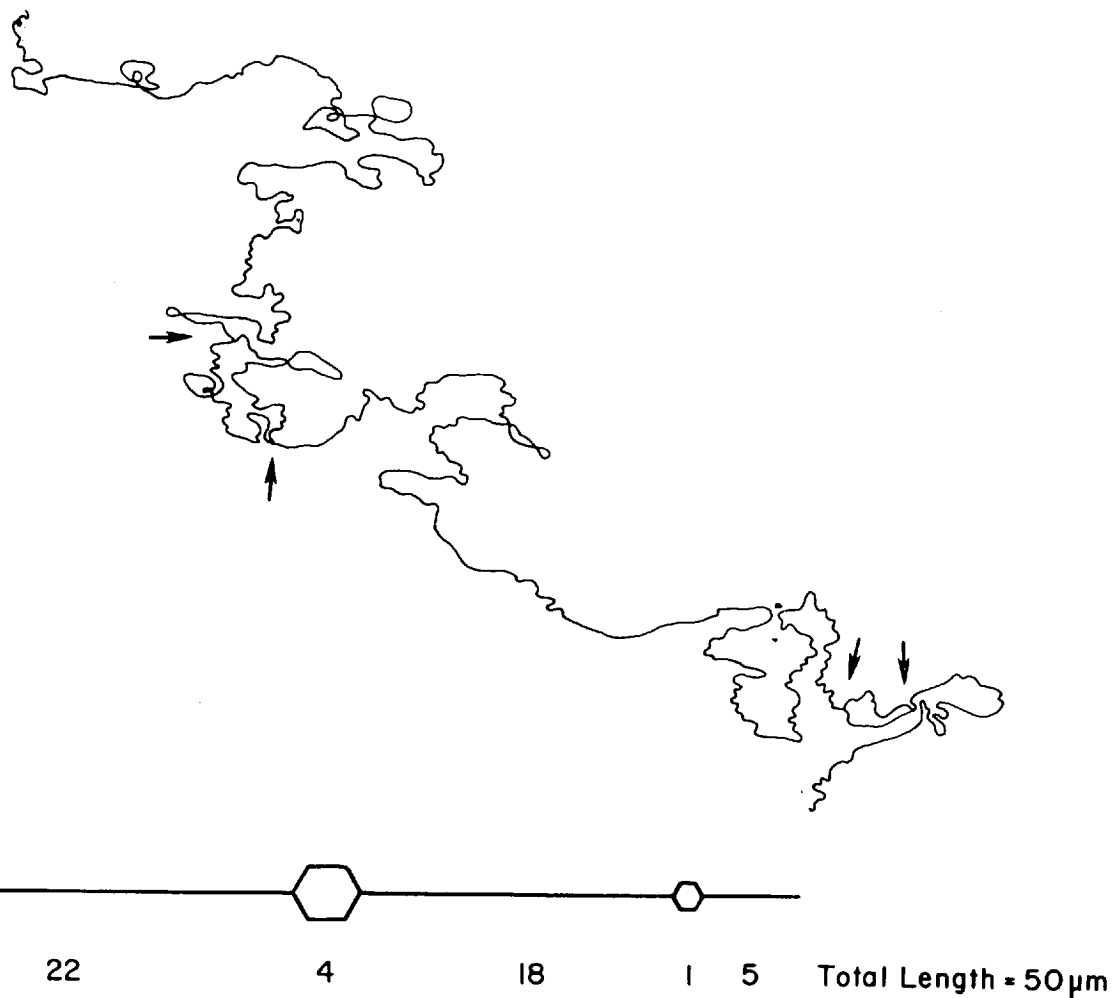


Figure 10. Tracing of an electron micrograph of a yeast DNA replication intermediate. DNA released from yeast spheroplasts was sedimented at 30,000 rpm for 9 hours. Material sedimenting at the position expected for T4-size DNA was examined by electron microscopy. The lower diagram gives the length of regions of the molecule in microns. The arrows indicate the positions of forks.

should be applicable to yeast replication intermediates and allow the mapping of replication start sites along a chromosomal DNA molecule. In addition, because yeast chromosomes are amenable to both genetic and molecular analyses, it should be possible to correlate sites on a DNA molecule with their corresponding genetic locations.

ACKNOWLEDGMENTS

This work was supported by a grant from the U.S. National Institutes of Health (GM-18926). W.B. was supported by a U.S. National Science Foundation predoctoral fellowship and by a U.S. Public Health Service training grant. C.S.N. was supported by an American Cancer Society postdoctoral fellowship. We thank Loretta Lenny for excellent technical assistance.

REFERENCES

- Blamire, J., D.R. Cryer, D.B. Finkelstein and J. Marmur. 1972. Sedimentation properties of yeast nuclear and mitochondrial DNA. *J. Mol. Biol.* 67: 11.
- Braun, R. and H. Wili. 1969. Time sequence of DNA replication in *Physarum*. *Biochim. Biophys. Acta* 174: 246.
- Bücking-Throm, E., W. Duntze, L.H. Hartwell and T.R. Manney. 1973. Reversible arrest of haploid yeast cells at the initiation of DNA synthesis by a diffusible sex factor. *Exp. Cell Res.* 76: 99.
- Chia, D. and V.N. Schumaker. 1974. A rotor speed dependent cross-over in sedimentation velocities of DNAs of different sizes. *Biochem. Biophys. Res. Comm.* 56: 241.
- Duntze, W., D. Stötzler, E. Bücking-Throm and S. Kalbitzer. 1973. Purification and partial characterization of α -factor, a mating type specific inhibitor of cell reproduction from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 35: 357.
- Finkelstein, D.B., J. Blamire and J. Marmur. 1972. Location of ribosomal RNA cistrons in yeast. *Nature New Biol.* 240: 279.
- Goldring, E.S., L.I. Grossman, D. Krupnick, D.R. Cryer and J. Marmur. 1970. The petite mutation in yeast. Loss of mitochondrial deoxyribonucleic acid during induction of petites with ethidium bromide. *J. Mol. Biol.* 52: 323.
- Hereford, L.M. and L.H. Hartwell. 1973. Role of protein synthesis in the replication of yeast DNA. *Nature New Biol.* 244: 129.
- Hohlfeld, R. and W. Vielmetter. 1973. Bidirectional growth of the *E. coli* chromosome. *Nature New Biol.* 242: 130.
- Mohburg, J. and H.P. Rusch. 1971. Isolation and DNA content of nuclei of *Physarum polycephalum*. *Exp. Cell Res.* 66: 305.
- Mortimer, R.K. and D.C. Hawthorne. 1969. Yeast genetics. In *The Yeasts* (A.H. Rose and J.S. Harrison, ed.) Vol. I, p.386. Academic Press, New York.
- , ———. 1973. Genetic mapping in *Saccharomyces*. IV. Mapping of temperature-sensitive genes and use of disomic strains in localizing genes. *Genetics* 74: 33.
- Mueller, G.C. and K. Kajiwara. 1966. Early- and late-replicating deoxyribonucleic acid complexes in HeLa nuclei. *Biochim. Biophys. Acta* 114: 108.

- Muldoon, J.J., T.E. Evans, O.F. Nygaard and H.H. Evans. 1971. Control of DNA replication by protein synthesis at defined times during the S period in *Physarum polycephalum*. *Biochim. Biophys. Acta* 247: 310.
- Newlon, C.S., T.D. Petes, L.M. Hereford and W.L. Fangman. 1974. Replication of yeast chromosomal DNA. *Nature* 247: 32.
- Petes, T.D. and W.L. Fangman. 1972. Sedimentation properties of yeast chromosomal DNA. *Proc. Nat. Acad. Sci. USA* 69: 1188.
- Petes, T.D., B. Byers and W.L. Fangman. 1973. Size and structure of yeast chromosomal DNA. *Proc. Nat. Acad. Sci. USA* 70: 3072.
- Petes, T.D., C.S. Newlon, B. Byers and W.L. Fangman. 1974. Yeast chromosomal DNA: Size, structure and replication. *Cold Spring Harbor Symp. Quant. Biol.* 38: 9.
- Schnös, M. and R.B. Inman. 1970. Position of branch points in replicating λ DNA. *J. Mol. Biol.* 51: 61.
- Slater, M.L. 1973. Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle. *J. Bacteriol.* 113: 263.
- Stubblefield, E. 1966. Mammalian chromosomes *in vitro*. XIX. Chromosomes of Don-C, a Chinese Hamster fibroblast strain with a part of autosome 1b translocated to the Y chromosome. *J. Nat. Cancer Inst.* 37: 799.
- Williamson, D.H. 1973. Replication of the nuclear genome in yeast does not require concomitant protein synthesis. *Biochem. Biophys. Res. Comm.* 52: 731.

Replication of polyoma DNA in isolated nuclei

Roger F. Martin

Nuclei isolated from polyomavirus infected 3T6 mouse fibroblasts support the continuation of replication of polyomavirus DNA when incubated *in vitro*. Polyomavirus can code for only 5 to 10 average size proteins, and thus utilizes largely host functions during replication, so that the system described constitutes a potential model for replication of DNA in mammalian cells. The *in vitro* DNA synthesis involves semi-conservative elongation of the daughter strands of the viral replicative intermediates, via the transient formation of short DNA chains which sediment at 4 to 5S in alkali. The DNA fragments are analogous to the Okazaki fragments in *Escherichia coli*, in that they contain RNA covalently attached to the 5' end. Fractionation of *in vitro* labelled replicating polyoma DNA by agarose gel electrophoresis, and subsequent analysis of the labelled DNA, suggests that the Okazaki fragments are associated with replicative intermediates at all stages of replication.

Polyomavirus is a small virus containing DNA in the form of a double stranded circle of approximately 5,000 base pairs, sufficient to code for 5 to 10 average size proteins. Infection of mouse cells with polyomavirus results in production of the virus and lysis of the infected cell. The replication of the viral DNA occurs in the cell nucleus and for polyomavirus, and its close relative SV40, it proceeds as shown diagrammatically in Figure 1. Replication is initiated at a unique origin, the two strands of the double helix separate and act as templates for the synthesis of the new (daughter) DNA strands. In the so-called replicative intermediates (RI), the daughter strands are elongated bi-directionally and the two replication forks finally meet at the termination point. The two parental strands, each hydrogen-bonded to a newly synthesised strand, then segregate to yield two double-stranded circular molecules [referred to as form II or Py(II)], each containing an intact (parental) circle and a daughter strand with a discontinuity at the termination point. Finally, the gap in the newly synthesised strand is filled-in and ligated, to yield mature form I polyoma DNA [Py(I)].

Nuclei from cells infected with polyomavirus contain viral DNA molecules at all stages of replication, but most of the viral

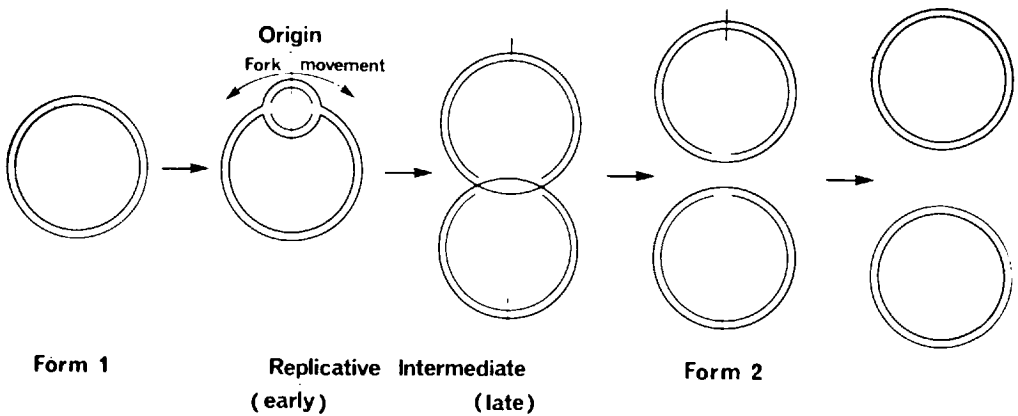


Figure 1. Replication cycle for polyoma and SV40 DNA

DNA (more than 95%) is as Py(I), with some Py(II). When nuclei isolated from infected cells are incubated *in vitro*, the growth of the polyoma DNA RI molecules continues semi-conservatively (Magnusson et al., 1972). In view of the limited capacity of the polyoma genome, most of the enzymes used for replication of the virus will be provided by the host, so that the investigation of the *in vitro* replication of the viral DNA is not only a study of polyoma DNA replication *per se*, but it also provides information on the biochemistry of cellular enzymes involved in the replication. Presumably the same enzymes are utilised for replication of cellular DNA.

METHODS

The sources of materials and the procedures used for infection of mouse 3T6 cells, isolation and incubation of nuclei, extraction of polyoma DNA, purification of polyoma replicative intermediates by BND cellulose chromatography, and analysis of the *in vitro* labelled DNA by centrifugation have been described previously (Magnusson et al., 1973; Winnacker et al., 1972). The conditions used for agarose gel electrophoresis were essentially the same as those described earlier (Tegtmeyer and Macasaet, 1972).

For preparative agarose gels, DNA was extracted from the gel slices by electrophoresis and concentrated by precipitation with ethanol. The author is indebted to Dr Gunnar Bjursell for the electron microscopy of the extracted DNA (Davis et al., 1971).

RESULTS AND DISCUSSION

Discontinuous elongation of daughter strands Isolated nuclei incorporate labelled deoxynucleoside triphosphates and Figure 2 shows the amount of α - ^{32}P -dGTP incorporated into acid insoluble viral DNA as a function of incubation time at 25°C. The rate of incorporation

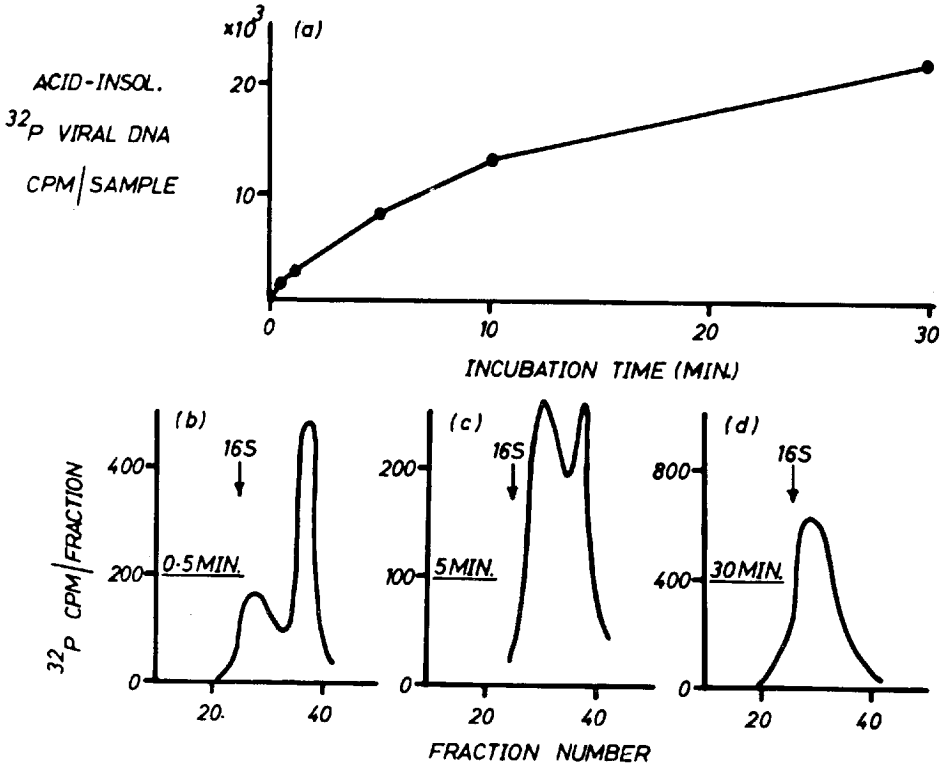


Figure 2. Incorporation of α - ^{32}P -dGTP into polyoma DNA in isolated nuclei. (a) Incorporation of ^{32}P into viral DNA (Hirt supernatant) as a function of incubation time. (b) - (d) Sedimentation analysis (sedimenting right to left) of ^{32}P -labelled viral DNA after 0.5, 5 and 30 min incubations.

decreases progressively, so that most viral DNA synthesis occurs before 30 min. Other data indicate that very little initiation [i.e. Py(I) to RI in Figure 1] occurs in the nuclei, and this could explain the limited capacity for incorporation. It is important to note that maturation is also suppressed in the isolated nuclei; only about 5% of the *in vitro* label is incorporated into Py(I), even after 30 min incubation. In a similar *in vitro* system for study of replication of polyoma DNA, employing a lysate of infected

mouse cells (Francke and Hunter, 1974), the capacity to initiate is also lost, but of the *in vitro* label incorporated into viral DNA, up to 50% is found as Py(I). This comparison suggests that the capacity to initiate DNA synthesis is a labile function which is inactivated on disruption of the cells. Also, it would seem that cytoplasmic protein(s) is involved in maturation of the polyoma DNA.

When the *in vitro* labelled DNA is analysed by alkaline sucrose gradient centrifugation, the labelled daughter strands dissociate from the template strands and sediment according to size. Full-size linear polyoma DNA sediments at 16S in alkali. Figures 2b to 2d show the results of alkaline sucrose gradient analysis of viral DNA extracted from nuclei incubated for 0.5, 5 and 30 min respectively. For short incubation periods, much of the *in vitro* label is incorporated into short pieces of DNA, but as the incubation time increases, the portion of label in short chains decreases. Moreover, there is a discontinuous size distribution consistent with the idea of a discontinuous mechanism of chain elongation, where the DNA is first synthesised in short pieces which are subsequently joined to the growing chain. The length of the intermediate DNA chains is approximately 100 to 200 nucleotides, as estimated from sedimentation data (Pigiet et al., 1974). Polyacrylamide gel electrophoresis should provide a more accurate length estimate. Discontinuous DNA synthesis was first observed in *Escherichia coli* by Okazaki (Okazaki et al., 1968) and the short intermediate DNA pieces are commonly referred to as Okazaki fragments. It should be noted that the Okazaki fragments observed in *E. coli* are considerably larger than those obtained from nuclei of polyoma infected cells.

The concept of discontinuous DNA synthesis arose from the consideration of the polarity of the DNA strands at the replication fork. All known DNA polymerases synthesize in the 5' to 3' direction, but although one of the growing strands at a replication fork can be elongated by simply adding nucleotides at the 3' end, the complementary daughter strand has the reverse polarity. This situation suggests that the growth of one strand might occur by synthesis of short intermediate chains (in the opposite direction to fork movement), and subsequent joining (Figure 3a). It is also possible that the synthesis is discontinuous on both strands, as shown in Figure 3b, but it is not yet completely clear which of the two alternatives operate in mammalian cells.

Like the Okazaki fragments from *E. coli* (Sugino et al., 1972) the short intermediate chains of polyoma DNA synthesised *in vitro* in the isolated nuclei system seem to contain a stretch of RNA attached to the 5' end of the DNA (Magnusson et al., 1973). Since the known DNA polymerases require a primer, this finding suggests

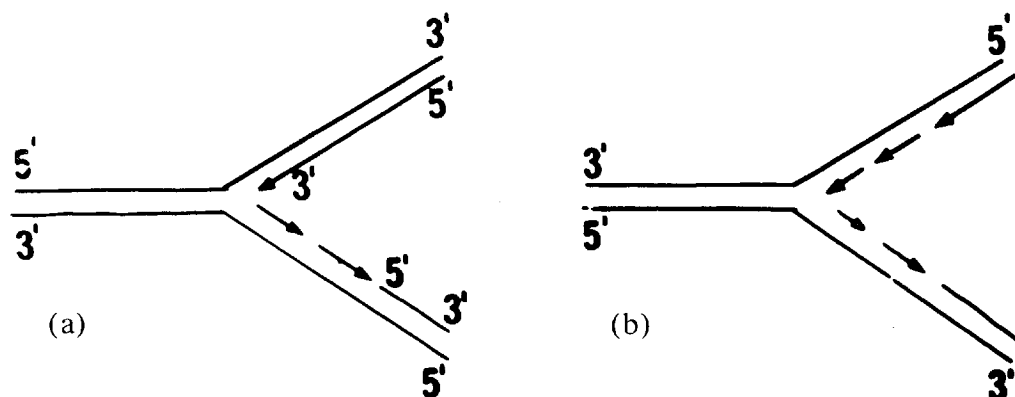


Figure 3. Models for (a) semi-discontinuous and (b) discontinuous replication of DNA

that the synthesis of Okazaki fragments is primed by a piece of RNA. If this is true one can speculate about mechanisms for the overall process of chain elongation. One possible model would involve RNA-primed synthesis on Okazaki fragments, removal of RNA primer, gap-filling and finally ligation.

Okazaki fragments at different stages of replication Agarose gel electrophoresis fractionates RI molecules according to maturity. Figure 4 shows the results of agarose gel electrophoresis of *in vitro* labelled RI, obtained from a 1.5 min incubation in the presence of α - ^{32}P -dCTP. The mobility of Py(I) and Py(II) in a parallel gel is also shown. When the gel fractions were pooled as indicated and the RI DNA extracted and analysed by sedimentation in alkaline sucrose gradients, each pool yielded a bi-modal distribution of ^{32}P -labelled DNA chains. The bi-modality is interpreted in the same way as the data in Figure 2; the lower molecular weight peak contains the Okazaki fragments and the other the larger growing daughter strands from the RI molecules. It is clear from Figure 1 that the length of the newly synthesised daughter strands is a direct reflection of the maturity of the RI molecules from which they were disassociated, so the decreasing sedimentation of the larger peak in the alkaline sucrose gradients A through D, indicates that the agarose gel did afford some fractionation of the RI according to maturity. (The general relationship between RI maturity and electrophoretic mobility was confirmed by extraction of DNA from agarose gel slices and examination by electron microscopy.) The presence of the 4-5S peak in all pools suggests that Okazaki fragments are associated with RI molecules at all stages of replication.

Superhelicity of replicative intermediates The simple diagrams in Figure 1 omit a very important structural feature of polyoma DNA -

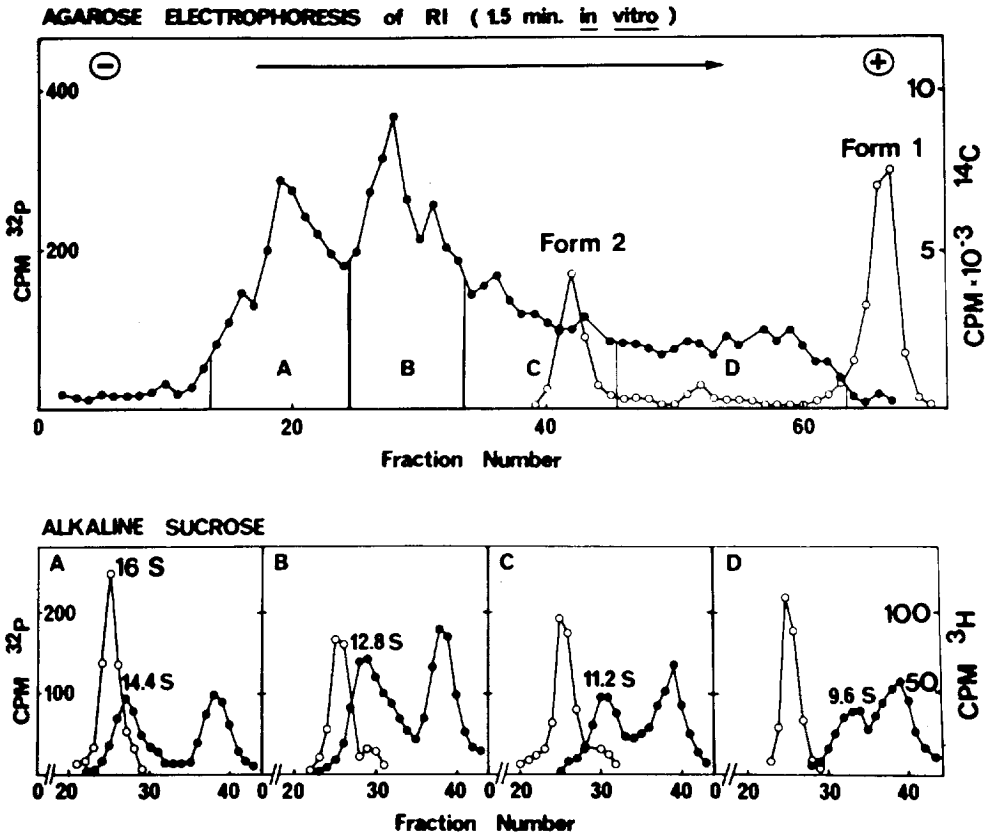


Figure 4. Sedimentation analysis of *in vitro* labelled daughter strands from replicative intermediates of different maturities. (Upper). Fractionation of replicative intermediates by agarose gel electrophoresis. (Lower). Sedimentation analysis (sedimenting right to left) of RI DNA extracted from agarose gel fractions pooled as indicated.

superhelicity. The duplex strand of Py(I), in addition to its own double helical structure, is further twisted to form a supercoiled structure (Figure 5). The superhelicity, a characteristic of circular duplex DNA molecules, results from a deficiency of turns in the duplex double helix. This deficiency, which is 'fixed' in circular molecules since they have no free ends, results in a 'tension' in a molecule which is partly relieved by supercoiling. However, if one strand of the duplex of Py(I) is broken, the tension in the molecule is relaxed as the degree of helicity in the duplex double helix returns to normal. Thus Py(II) can be obtained from Py(I) by simply introducing a single-stranded break, for example by 'nicking' with a low concentration of DNAase.

The non-replicated region of RI molecules is also supercoiled, as shown in Figure 5. The structures of Figure 5d are actually

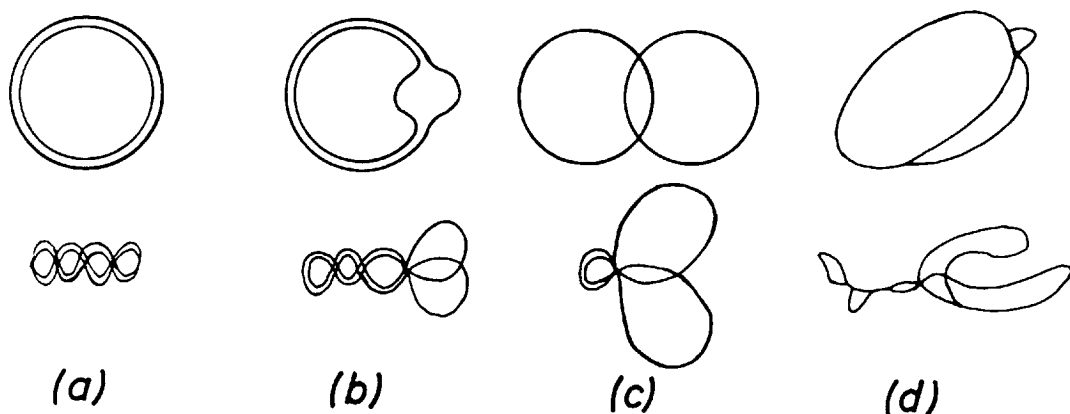


Figure 5. Relaxed and supercoiled forms of polyoma DNA. (a) Mature polyoma DNA, Form I (supercoiled) and Form II (relaxed). (b) and (c) Relaxed and supercoiled forms of early and late replicative intermediates, respectively. (d) Tracings of molecules from electron micrographs of RI DNA extracted from a preparative agarose gel.

tracings from molecules found in electron micrographs of RI DNA extracted from preparative agarose gels. In view of the marked difference in the electrophoretic mobilities of Py(I) and Py(II), it would be expected that superhelicity would also influence the mobility of RI molecules. This was confirmed by the results of the experiment shown in Figure 6. ^{32}P -labelled RI DNA was extracted from a few fractions from a preparative agarose gel, mixed with ^{14}C -Py(I) and re-electrophoresed on agarose gels either before or after treatment with a low concentration of DNAase. In the control, most of the ^{32}P -RI DNA re-electrophoresed with the expected mobility. After DNAase treatment, the ^{14}C -Py(I) was converted to Py(II) and part of the ^{32}P -RI migrated with a considerably reduced mobility while the remainder of the ^{32}P -RI had the same mobility as in the control. This result suggested that the RI preparation was a mixture of both relaxed and supercoiled molecules (the latter being more mature so that all RI molecules had about the same electrophoretic mobility), and that DNAase treatment relaxed the supercoiled RI molecules, yielding RI having a lower electrophoretic mobility.

Some of the relaxed RI molecules undoubtedly arose by 'nicking' of intact molecules during extraction of the viral DNA from the nuclei, or during the *in vitro* incubation, since some of the mature polyoma DNA is extracted as Py(II) rather than Py(I). It is also possible that some relaxed RI molecules are real intermediates in

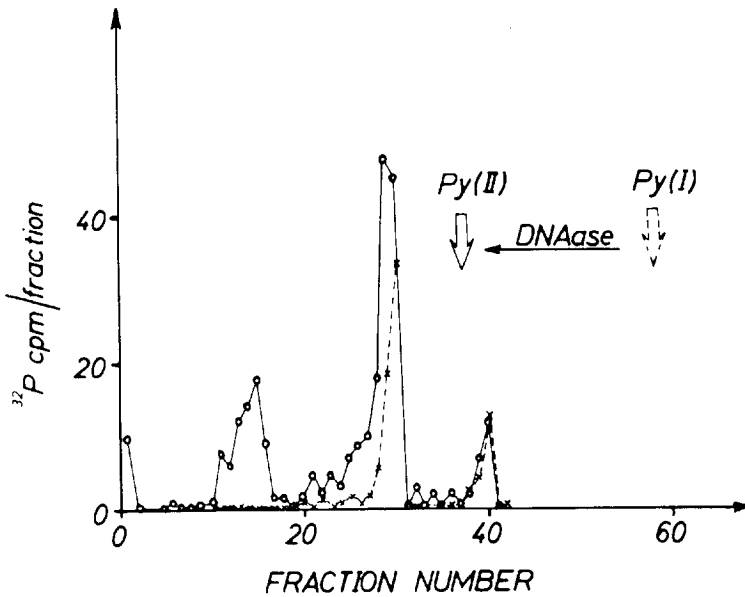


Figure 6. Effect of DNAase treatment on the electrophoretic mobility (direction of electrophoresis, left to right) of RI molecules. DNAase treated (○—○), control (x - - - x). The arrows show the positions of ^{14}C - Py(I) and Py(II).

the replication process. Movement of the replication fork requires unwinding of the double helix and in polyoma RI molecules this would require nicking of at least one of the parental strands, presumably followed quickly by rejoining, and in fact a protein with this kind of endonuclease-ligase activity has already been described (Champoux and Dulbecco, 1972).

A more practical consequence of the presence of both relaxed and supercoiled RI molecules is that it complicates the fractionation of RIs by agarose gel electrophoresis; there is no simple relationship between mobility and maturity. Figure 7 summarises the results of a series of experiments similar to that shown in Figure 6. The maturity of the RI molecules was estimated by sedimentation analysis (Studier, 1965) of the labelled daughter strands. In one experiment relaxed and supercoiled RIs were partly separated by electrophoresis in agarose gels containing ethidium bromide (Sharp et al., 1973). The important conclusion to be obtained from the results in Figure 7 is that agarose gel electrophoresis allows preparation of RIs of relatively defined maturity only for supercoiled RIs up to about 40% replicated or relaxed RIs greater than about 70% replicated.

Preliminary experiments indicate that fractionation of RIs by agarose gel electrophoresis and subsequent analysis of the Okazaki

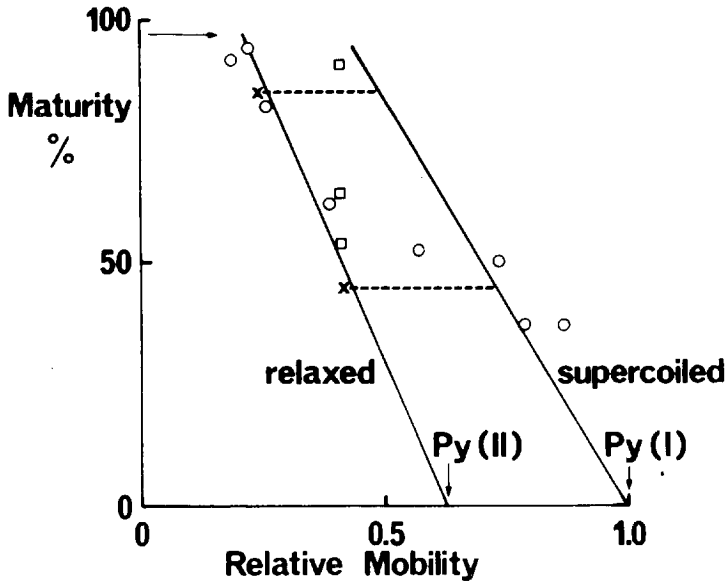


Figure 7. Approximate relationship between electrophoretic mobility (relative to Py(I)) and maturity of both relaxed and supercoiled RI molecules. Direct sedimentation analysis of RI DNA extracted from 1 or 2 fractions of a preparative agarose gel (○). Sedimentation analysis after ethidium bromide agarose gel electrophoresis (□). Effect of DNAase on mobility (x).

fragments by electrophoresis in formamid-polyacrylamide gels, similar to those described previously (Maniatis and Ptashne, 1973), enables comparison of the size distribution of Okazaki fragments from RIs of different maturities.

CONCLUSIONS

The study of polyoma replication in isolated nuclei provides an *in vitro* system for investigation of the biochemistry of DNA synthesis in the mammalian cell. It might be expected that further studies will yield details of the number and properties of the enzymes involved in the replication process. However, even the information obtained in this stage provides a useful base for investigation of the biochemistry of the replication of mammalian DNA.

ACKNOWLEDGMENTS

The experiments detailed in this paper were carried out by the author in the Department of Biochemistry, Karolinska Institute, Stockholm, during the course of a post-doctoral fellowship supported by the Cancer Institute, Melbourne. The author is indebted to Professor P. Reichard and Drs Gunnar Bjursell, Goran Magnusson, Bernd Otto and Vicent Pigiet for valuable discussions.

REFERENCES

- Champoux, J.C. and R. Dulbecco. 1972. An activity from mammalian cells that untwists superhelical DNA - a possible swivel for DNA replication. *Proc. Nat. Acad. Sci. USA* 69: 143.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *In Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol.21D, p.413. Academic Press, New York.
- Francke, B. and T. Hunter. 1974. *In vitro* polyoma DNA synthesis: Discontinuous chain growth. *J. Mol. Biol.* 83: 99.
- Maniatis, T. and M. Ptashne. 1973. Multiple repressor binding at operators in bacteriophage λ . *Proc. Nat. Acad. Sci. USA* 70: 1531.
- Magnusson, G., E.L. Winnacker, R. Eliasson and P. Reichard. 1972. Replication of polyoma DNA in isolated nuclei. II. Evidence for semi-conservative replication. *J. Mol. Biol.* 72: 539.
- Magnusson, G., V. Pigiet, E.L. Winnacker, R. Abrams and P. Reichard. 1973. RNA-linked short DNA fragments during polyoma replication. *Proc. Nat. Acad. Sci. USA* 70: 412.
- Okazaki, R., T. Okazaki, K. Sakabe, K. Sagimoto, R. Kainuma, A. Sugino and N. Iwatsuki. 1968. *In vivo* mechanism of DNA chain growth. *Cold Spring Harbor Symp. Quant. Biol.* 33: 129.
- Pigiet, V., R. Eliasson and P. Reichard. 1974. Replication of polyoma DNA in isolated nuclei. III. The nucleotide sequence of the RNA-DNA junction in nascent strands. *J. Mol. Biol.* 84: 197.
- Sharp, P.A., B. Sugden and J. Sambrook. 1973. Detection of two restriction endonuclease activities in *H. parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* 12: 3055.
- Studier, F.W. 1965. Sedimentation studies of the size and shape of DNA. *J. Mol. Biol.* 11: 373.
- Sugino, A., S. Hirose and R. Okazaki. 1972. RNA-linked nascent DNA fragments in *E. coli*. *Proc. Nat. Acad. Sci. USA* 69: 1863.
- Tegtmeyer, P. and F. Macasaet. 1972. SV40 DNA synthesis: Analysis by gel electrophoresis. *J. Virol.* 10: 599.
- Winnacker, E.L., G. Magnusson and P. Reichard. 1972. Replication of polyoma DNA in isolated nuclei. I. Characterisation of the system from mouse fibroblast 3T6 cells. *J. Mol. Biol.* 72: 523.

Chromosome functional organization determined by replication mapping

Chev Kidson, Phillip C. Chen, Tony Edwards and Jonathan Mansbridge

The thymidine analog, 2', 3'-dideoxythymidine (ddT) is rapidly phosphorylated and incorporated into DNA in replicating mammalian cells where it terminates growing daughter chains. In synchronized cell populations apparent mutations are induced by ddT with high frequency and with maxima which are localized in time. Pulsed chain termination can thus be employed to produce a replication map. Coupled with chromosome allocation of genes by cell fusion techniques this approach permits in principle the drawing of replication maps of individual chromosomes. The data obtained are consistent with the existence in diploid and heteroploid cells of a highly ordered replication program. So far there are indications that certain functionally co-ordinated gene loci are closely co-ordinated in the replication map, e.g. cistrons coding for structurally distinct subunits of a polymeric enzyme which lie on different chromosomes; loci for enzymes which are contiguous in a metabolic pathway but lie on different chromosomes. The paradoxically low lethality observed is difficult to correlate with high mutation frequencies. Chromosome structural anomalies are seen but these are less severe and gross breaks are less frequent than expected. Preliminary evidence for DNA sequence restriction in ddT incorporation might resolve the paradox in part and also suggests ways in which chain terminators might be further used to probe chromosome structure.

The advent of cell fusion techniques has led to the assignment of gene loci to particular mammalian chromosomes, especially in human cells (Ruddle, 1973). In certain cases it is possible to allocate gene functions to subregions of individual chromosomes by coupling information derived from naturally occurring translocations and deletions with segregation analysis following cell fusion (Ricciuti and Ruddle, 1973). It is conceivable that logical extension of these approaches will eventually permit the construction of mammalian chromosome maps in a rough topological sense: The difficulty of relating such maps to linear DNA order within chromosomes will presumably remain, in view of the uncertainties about the folding of chromosomal DNA. In contrast to the situation in *Drosophila*, there is currently no satisfactory detailed explanation even for such crude parameters of mammalian chromosome structure as Giemsa banding. In this sense, then, it will be some time before meaningful interpretation of crude

topological mammalian chromosome maps becomes possible. It is difficult at present to foresee the acquisition of detailed knowledge of gene positions in relation to chromosome organization such as has been gained for bacterial chromosomes by recombination analysis, except perhaps around integration sites of transforming viral genomes. In bacteria somewhat similar information has also been provided by replication mapping with N-methyl-N'-nitro-N-nitrosoguanidine (Cerda-Olmedo and Hanawalt, 1968) but unfortunately in mammalian cells this mutagen does not show replication point preference (Orkin and Littlefield, 1971).

Logically, DNA chain-terminating nucleosides should act as replication point mutagens, with the expectation that they would give rise to deletion mutations. From this premise, we have examined the thymidine analog, 2',3'-dideoxythymidine (ddT), as a potential tool for mapping mammalian chromosomes. The success of such an approach must depend on rapid phosphorylation and polymerization of the analog, together with inability of DNA repair systems to correct the damage completely. The triphosphate of this nucleoside can be used as a substrate by DNA polymerase I of *Escherichia coli* and incorporated into growing DNA chains *in vitro*, whereupon the chains terminate because of the lack of the 3'-hydroxyl group necessary for chain continuation. However, the nucleoside itself is a poor substrate for *E. coli* thymidine kinase (Atkinson et al., 1969). Fortunately both kinase and replication enzymes can handle the analog well in mammalian cells (Byars and Kidson, 1974).

This enables us to use the chain terminator as a probe of DNA replication and so to tackle a number of questions: How precise is the replication order of genetic markers on mammalian chromosomes? Is there co-ordination of the replication of allelic loci and thus of homologous chromosome pairs? Are daughter strands at a given locus replicated precisely together? Does heteroploidy alter the program? Is there any functional significance of a replication program? In addition, nucleotide sequence considerations suggest that chain terminators may be useful probes of chromosome structure *per se*.

DNA CHAIN TERMINATION IN MAMMALIAN CELLS

All mammalian cell lines examined phosphorylate ddT and incorporate the corresponding nucleotides into DNA rapidly; the rate and extent varies with the cell line. Initial chemical studies were carried out with the mouse myeloma XCl, since this was found to have the highest incorporation of ddT into DNA (Byars and Kidson, 1974). A concentration of 1 mM ddT is required to give 95% inhibition of incorporation of 0.2 μ M 3 H-thymidine in

these cells, which suggests that ddT incorporation may not involve simple competition with thymidine. Alkaline sucrose density gradient analysis shows that ddT-labelled daughter strands are mostly shorter than parental strands. A substantial proportion (10-15%) of incorporated ddT appears in oligonucleotides of 2-20 residues long, as judged from ion-exchange chromatography and polyacrylamide gel electrophoresis; some of these oligomers may be palindromic, as judged by their elution from hydroxylapatite at high ionic strength. Standard enzymatic analysis shows that in both long and short chains ddT is at the 3'-end. This renders the chains relatively resistant to exonuclease I, a finding which suggests that they could be somewhat resistant to exonuclease attack *in vivo*.

Oligonucleotides appearing after early *S* phase pulsing of cells with ³H-ddT are not present late in the same *S* phase. Some of the label incorporated into longer chains is also lost during this *S* phase, but some ddT remains in DNA over periods equivalent to two further cell cycles (Byars and Kidson, 1974). These data indicate that some repair occurs, but that some terminated chains persist for a considerable time. The nature of the repair is not yet known, especially whether it involves double strand breaks, a question of importance in relation to the use of ddT as a mutagen and as a probe of chromosome structure. However, the fact that ddT is rapidly incorporated as a DNA chain terminator in mammalian cells and that some terminated chains persist suggests that this nucleoside should be a replication point mutagen and that the mutations should include deletions. The observation that cell growth is slowed in proportion to the time of ddT exposure (Figure 1) but that many cells survive treatment and ultimately attain normal growth rates, has encouraged us to explore the potential of this compound as a mutagen which might give information on mammalian chromosome functional organization.

REPLICATION MAPPING OF CHROMOSOMES IN HUMAN DIPLOID CELLS

The pulsed application of ddT during *S* phase in synchronized cells, followed by agar drop cloning (Patuleia and Friend, 1967) and histochemical or autoradiographic analysis of clones is the approach which has been used in replication mapping (Chen and Kidson, unpublished) (Figure 2). The choice of male human diploid cells (HEF 1) for initial studies was deliberate; the single X chromosome provides a set of haploid genetic markers. Chromosome assignments have been made by cell fusion methods for many human chromosomes (Ruddle, 1973). The markers used here and their chromosome assignments are listed in Table 1. Hourly pulses of ddT in the 10 hour *S* phase of HEF 1 distinguishes mutation peaks

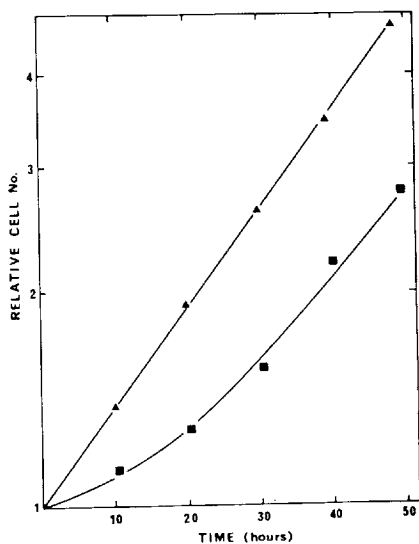


Figure 1. Growth curves of mouse myeloma cells pulsed with ddT for 1 hour and untreated controls

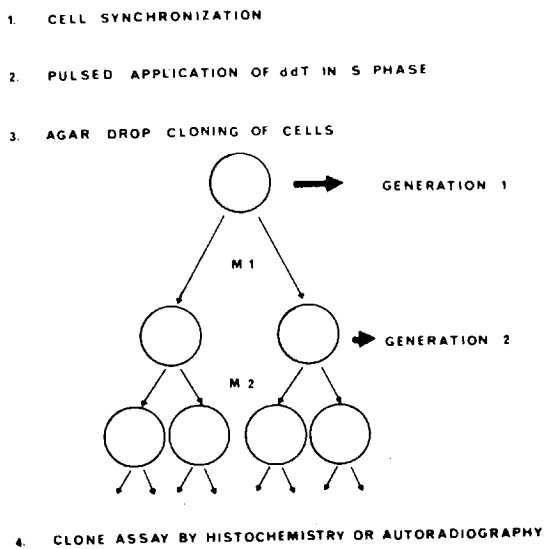


Figure 2. Experimental design of replication mapping experiments

for LDH (hour 2) and AcP (hour 6) from those of G6PDH, 6PGDH and HGPRT (hour 5). The latter group can be separated by 10 min pulse times, so that these five markers can be placed on a time map (Figure 3).

TABLE 1

Enzyme markers and their chromosome assignments

<u>Marker</u>		<u>Chromosome no.</u>
Lactate dehydrogenase (LDH)	A subunit	11
	B subunit	12
Hypoxanthine-guanine phosphoribosyl transferase	(HGPRT)	X
Glucose-6-phosphate dehydrogenase	(G6PDH)	X
6-phosphogluconate dehydrogenase	(6PGDH)	1
Acid phosphatase	(AcP)	2

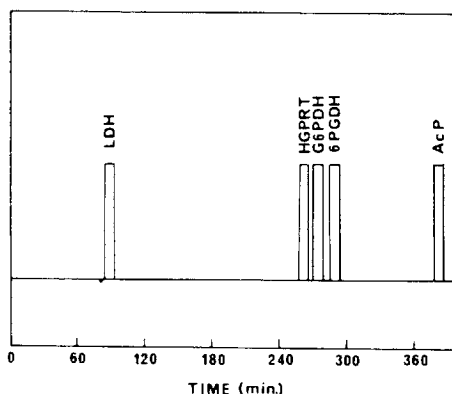


Figure 3. Replication map of five markers in HEF 1 cells, based on 10 min ddT pulse data

G6PDH and HGPRT are X-linked markers and thus have only haploid representation in the male genome of HEF 1. The other three markers are diploid. The finding of localized mutation maxima for all five markers is consistent with the occurrence of mutations at DNA replication points (cf. Cerda-Olmedo and Hanawalt, 1968) and suggests that in the case of diploid loci,

alleles are hit close together in time by the mutagen. Now, these data are based on cloning carried out on first generation cells, before the first mitosis. When cells were allowed to pass through one mitosis before cloning, the observed mutation frequencies were very similar. The finding that there is little or no segregation in the second generation suggests that ddT mutagenesis affects both daughter strands at a given locus in a chromosome. This could mean either that both are hit simultaneously or that closely co-ordinated replication of both daughters is inhibited when one is terminated by ddT. However, some caution must be entertained here; segregation in the second or subsequent generations would depend on the genesis of mixed clones in the first generation. Fibroblast clones are small and although some mixed clones undoubtedly occur, they are usually very difficult to distinguish with certainty, depending on the topological distribution of positive and negative cells within a given clone.

FUNCTIONALLY LINKED MARKERS

The close relationship in time of the replication positions of G6PDH and 6PGDH is interesting in view of their functional association in the pentose phosphate pathway and their different chromosomal allocations. The possibility that this could reflect a control coupled to replication is attractive; the observation is at least a stimulus to attempt mapping of other metabolically related, chromosomally distinct enzyme markers.

The single map position of LDH is intriguing in view of the fact that the two LDH subunits are coded by genes on different chromosomes. This suggests either that both genes are replicated close together in time (i.e. a second example of functional co-ordination) or that mutation at one locus, e.g. LDH-A, leads to failure of expression of the second, LDH-B, even though theoretically the cell might be expected still to have the potential to synthesize LDH-5 (B4). In this respect it is of interest that HEF 1 control cells produce little or no LDH-5. The system merits further analysis using isolated mutants.

CHROMOSOME REPLICATION PATTERNS

The map position of the earliest marker so far observed, LDH, occurs before any readily measurable rise in ^3H -thymidine incorporation into DNA. This is consistent, however, with the early replication of chromosome 11 (Breg et al., 1972). The kinetics of ^3H -thymidine incorporation may simply reflect initial residual pools of unlabelled thymidine nucleotides referable to the thymidine block used for synchronization (Chen and Kidson, unpublished). Replication of chromosome 1 begins early in S

phase (Slezinger and Prokofieva-Belgovskaya, 1968) before the commencement of replication of chromosome 11, so that the position of the 6PGDH marker must be some distance along the replication map for that chromosome, since mutations appear in the fifth hour of *S* phase. The position of AcP must be even further along the replication map for chromosome 2, since mutations occur in the sixth hour and replication commences early in *S* phase (Slezinger and Prokofieva-Belgovskaya, 1968). The X-linked markers have variably been reported as being on the long arm (Ricciuti and Ruddle, 1973) or on the short arm (Grzeschik et al., 1972) of the X chromosome. Variable estimates have been made of the closeness of their linkage but the reported separation of the two loci in cell hybrids (Miller et al., 1971) suggests they may be some distance apart. The fact that their replication order can be distinguished suggests that it should be possible eventually to construct a more complete replication map of the X chromosome. This should also be true for autosomes but will depend on further chromosome allocation of suitable markers by cell fusion analysis.

MAPPING OF TUMOR CELL GENOMES

There are suggestions from chromosome labelling experiments that diploid cells from different tissues in the same animal may have somewhat different replication patterns (Slezinger and Prokofieva-Belgovskaya, 1968). This clearly is important in terms of the potential general usefulness of replication mapping as a descriptive parameter of chromosome organization. Conceivably tumor cells, especially where these are heteroploid, could exhibit serious variation. However, HeLa cells give a map for four markers very similar to that for HEF 1, when map units are expressed as fractions of *S* phase (Figure 4). Although there is some difficulty in delineating the precise end point of *S* phase, the relative order of marker map positions is clearly similar. There are, however, suggestions of segregation for G6PDH and 6PGDH after second generation cloning in HeLa even though the relevant chromosomes (X and 1) are diploid in this clone. Preliminary data suggest that relative map positions in Raji cells (transformed diploid) are similar to HEF 1 and HeLa. Mouse neuroblastoma cells have a somewhat different map from the human lines but have not yet been compared with diploid mouse cells. At this stage the data seem promising enough to suggest that replication mapping with ddT will prove to be of general use in mammalian and perhaps other eukaryotic cell lines.

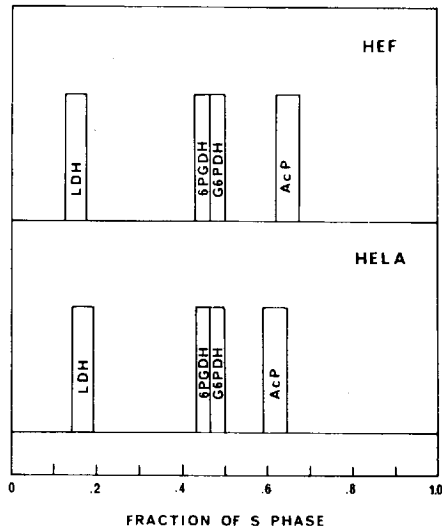


Figure 4. Normalized replication maps of HEF 1 and HeLa cells, based on hourly ddT pulse data

MUTATION FREQUENCIES AND LETHALITY

Apparent mutation frequencies range from about 0.10 to 0.35 (Chen and Kidson, unpublished), which are exceptionally high for any mutagen. If we assume that clones scored as negative are indeed negative clones and not mixed clones, then observed mutation frequencies would give a measure of 4-hit events, i.e. both daughter strands of both chromatids at a particular diploid locus. If we correct for asynchrony in the cell population, the above frequency range becomes 0.14 to 0.50, figures which would require single hits to occur with frequencies of 0.61 to 0.84. Not only would truly negative clones seen on first generation cloning reflect 4-hit events, hits must involve the parental as well as the daughter strand in each case, so that no effective gene product is produced by that locus after subsequent DNA replication cycles. This would eliminate repair of a type which involved simple excision of ddT and gap-filling using the parental strand template, with errors. More probably, such high mutant frequencies might be expected to result from *effective double strand hits*, i.e. parental strand breaks opposite daughter strand gaps, giving rise to actual double strand breaks. Conceivably these might remain as such (cf. radiochemical evidence of persisting ddT) or might be followed by a degree of intrastrand recombination.

Mixed clones are observed and there is probably some segregation in HeLa cells. These should reflect asymmetric 2-hit, or 3-hit events. The difficulty in estimating their frequencies, as explained above, lies in identification of mixed clones.

It would be expected that mutation frequencies of this order would be accompanied by very high lethality. Surprisingly this is not so (Figure 5). While estimates of survival in HEF 1 might be substantially in error due to the very low cloning efficiency of primary fibroblasts, mouse neuroblastoma cells have a cloning efficiency of about 90%, yet give very low lethality. In both cases maximum lethality occurred early in *S* phase.

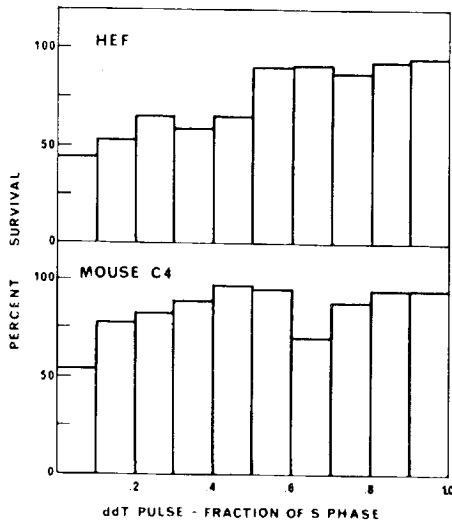


Figure 5. Clone survival in HEF 1 and mouse neuroblastoma cells, expressed as a percentage of surviving controls following pulsed application of ddT in synchronous *S* phase

These observations provide us with a paradox. If the high mutation rates can be explained by effective double strand, 4-hit events due to a high degree of replicative synchrony these very events should be lethal. If breaks remain unjoined, the chromosomal DNA should fragment and we should expect to see shattered chromosomes. If breaks are rejoined, the number of deletions should be high enough to hit a large number of essential functions

and so be lethal. Either way lethality should be high but surprisingly is not. In the expectation that some clues to this paradox must be readily available we have begun to examine chromosome structure following ddT exposure and DNA sequences into which ddT is incorporated.

CHROMOSOME STRUCTURE

Chromosomes have been examined in the first metaphase following whole *S* phase or hourly pulsed exposure to ddT in synchronized HEF 1. Although quantitative data are not yet available some general statements can be made (Table 2). The most important observation is that Giemsa staining with or without trypsinization does not reveal shattered chromosomes, nor even a high frequency of chromatid breaks. The latter do occur, are more frequent following whole *S* phase exposure and usually affect only one chromatid of a homologous chromosome pair. On the other hand, a variety of less well-defined anomalies is observed, especially abnormal banding patterns, which could reflect the existence of deletions but which have not been specifically defined at this time. Some anomalies observed could conceivably reflect mutations in chromosome structural protein genes.

TABLE 2
Chromosome structure after ddT exposure

	Whole <i>S</i> phase exposure	Hourly pulsed exposure
Major chromatid breaks	Moderate	Few
Giemsa staining	'Lumpy'	Normal
Trypsin banding	Some loss of pattern	Mostly normal pattern
Metaphase chromatid separation	Not seen	Occasional
Abnormal chromosome number	Occasional	Not seen

The failure to find widespread breaks even following trypsinization is consistent with the idea that in cells which can progress as far as metaphase after being hit by ddT, DNA double strand breaks may be largely repaired. Alternatively

DNA fragments do exist but are held together by proteins which are not sensitive to short term trypsin digestion.

DNA SEQUENCE CONSIDERATIONS

If chromosomal integrity is maintained in a general sense, this is at least consistent with high cell survival rates, even if we do not yet know the detailed chemistry involved in maintaining structure. Low lethality must also reflect some degree of avoidance of hit by ddT of many essential functions. If there were some sequence selectivity in incorporation of ddT into DNA it would conceivably explain the low lethality. Preliminary examination of the possibility is encouraging, namely, there is a markedly disproportionate incorporation of ddT compared with thymidine into longer pyrimidine tracts *vis a vis* monomers and short pyrimidine tracts. This suggests sequence-restricted incorporation, although we do not yet know the extent of the constraints. The possibility that pyrimidine tracts, especially thymine tracts, may occur at particular positions and only in some replicons can be entertained. So too can the possibility that they have particular roles in relation to the packaging of chromosomal DNA and/or in replication or transcription. In addition to ddT, other chain-terminating nucleosides such as dda and ddC should be useful in giving complementary information.

What will be interesting to discover is whether deletion occurs of part of a pyrimidine tract or of a very short sequence containing such a tract (such as might be envisaged if this were near the end of a replicon) or whether deletion occurs of a much larger segment (such as might be envisaged if the sequence hit were near the beginning of a replicon). If the latter is the case, it would also be useful to know if there is any relation at all between deleted sequences, repetitive models of chromosome sequence arrangements (Davidson et al., 1973; Pyeritz and Thomas, 1973) and functional chromosome organization.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Health and Medical Research Council, the Apex Foundation for Mental Retardation and the Queensland Cancer Fund.

REFERENCES

- Atkinson, J.R., M.P. Deutscher, A. Kornberg, A.F. Russell and J.G. Moffat. 1969. Enzymatic synthesis of deoxyribonucleic acid. XXXIV. Termination of chain growth by a 2', 3'-dideoxyribonucleotide. *Biochemistry* 8: 4897.

- Breg, W.R., P.W. Alderdice, D.A. Miller and O.J. Miller. 1972. Quinacrine fluorescence patterns and terminal DNA labelling of human C group chromosomes. *Nature New Biol.* 236: 76.
- Byars, N. and C. Kidson. 1974. DNA chain termination by 2',3'-dideoxythymidine in replicating mammalian cells. *Biochemistry*. In press.
- Cerda-Olmedo, E. and P. Hanawalt. 1968. The replication of the *Escherichia coli* chromosome studied by sequential nitrosoguanidine mutagenesis. *Cold Spring Harbor Symp. Quant. Biol.* 33: 599.
- Davidson, E.H., B.R. Hough, C.S. Amenson and R.J. Britten. 1973. General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* 77: 1.
- Grzeschik, K.H., P.W. Alderdice, A. Grzeschik, J.M. Opitz, O.J. Miller and M. Siniscalco. 1972. Cytological mapping of human X-linked genes by use of somatic cell hybrids involving an X-autosome translocation. *Proc. Nat. Acad. Sci. USA* 69: 69.
- Miller, O.J., P.R. Cook, P.M. Khan, S. Shin and M. Siniscalco. 1971. Mitotic separation of two human X-linked genes in man-mouse somatic cell hybrids. *Proc. Nat. Acad. Sci. USA* 68: 116.
- Orkin, S. and J.W. Littlefield. 1971. Nitrosoguanidine mutagenesis in synchronized hamster cells. *Expt. Cell Res.* 66: 69.
- Patuleia, M.C. and C. Friend. 1967. Tissue culture studies on murine virus-induced leukemia cells: Isolation of the single cells in agar-liquid medium. *Cancer Res.* 27: 726.
- Pyeritz, R.E. and C.A. Thomas. 1973. Regional organization of eukaryotic DNA sequences as studied by the formation of folded rings. *J. Mol. Biol.* 77: 57.
- Ricciuti, F. and F.H. Ruddle. 1973. Assignment of nucleoside phosphorylase to D-14 and localization of X-linked loci in man by somatic cell genetics. *Nature New Biol.* 241: 180.
- Ruddle, F.H. 1973. Linkage analysis in man by somatic cell genetics. *Nature* 242: 165.
- Slezinger, S.I. and A.A. Prokofieva-Belgovskaya. 1968. Replication of human chromosomes in primary cultures of embryonic fibroblasts. *Cytogenetics* 7: 337.

Control of DNA synthesis in somatic cell hybrids

Jennifer A. Marshall Graves

The hamster chromosome set in mouse x hamster cell hybrids retains its original rate and sequence of replication, even in clones which have lost different sets of hamster chromosomes. These results favour the hypothesis that DNA synthesis in chromosomes (or smaller units) is controlled autonomously in mammalian cells.

Chromosomal DNA synthesis appears to follow a heritable and well-defined schedule in adult mammalian cells. Density labeling and autoradiographic experiments have shown that the sequence of replication is consistent from cycle to cycle (Mueller and Kajiwara, 1966) and from cell to cell, even in different tissues, different animals and different species (Martin, 1966; Graves, 1967). These observations indicate that the course of DNA synthesis is regulated very precisely during the S phase, but at present the nature of this regulation remains quite unclear.

Control of the different phases of the cell cycle has been examined by the techniques of cell fusion and cell hybridization. Studies with fused cells have demonstrated that cytoplasmically transmissible factors control the initiation of DNA synthesis and the onset of mitosis (Rao and Johnson, 1970; Graves, 1972a). However, between these two points of synchronization, the two chromosome sets in heterokaryons appear to follow their original schedule of DNA synthesis autonomously (Graves, 1972a, 1972b). Similarly in hybrids between mouse cells (having a long S phase) and hamster cells (with a short S), the chromosome sets began DNA synthesis synchronously, but the hamster set terminated considerably earlier than the mouse set, apparently in the same sequence as in the hamster parent.

To examine further the question of autonomy of control, the rate and pattern of chromosome replication has been compared in a series of reduced mouse x hamster cell hybrids from which different overlapping sets of hamster chromosomes have been lost. If replication is controlled autonomously in individual chromosomes or smaller units, the duration and sequence of DNA synthesis in the rest of the

complement would be expected to remain unchanged. However, if the synthesis schedule for the hamster complement as a whole is controlled by factors located on one or a few chromosomes, absence of particular hamster chromosomes from reduced hybrids could cause modification of the timing of synthesis in the remainder of the complement. It might then be possible to assign such factors to particular chromosomes.

MATERIALS AND METHODS

Cell lines and culture methods The B₁₄I₅₀ line of BUdR-resistant Chinese hamster cells (obtained from the American Type Culture Collection) has low thymidine kinase activity, and is therefore unable to grow in HAT medium. NCTC 2472 is a permanent mouse line, and NCTC (az^r), an azaguanine-resistant variant of it, also are unable to grow in HAT medium. Cells were grown as monolayers in Dulbecco Modified Eagle's medium (10% fetal calf serum) containing 50 units/ml of penicillin and streptomycin. Hybrids between BI₅₀ and NCTC (az^r) were selected in HAT medium, containing 10⁻⁴ M hypoxanthine, 4 x 10⁻⁷ M aminopterin and 1.6 x 10⁻⁵ M thymidine (Littlefield, 1964).

Cell hybridization 2:1 mixtures of NCTC (az^r) and BI₅₀ cells were fused with inactivated Sendai virus as described previously (Graves, 1972a), and plated in HAT medium. Colonies were isolated, propagated and karyotyped and hybrid clones were identified by the presence of parental marker chromosomes. A hybrid between BI₅₀ and NCTC, obtained previously by a half-selective method, was also studied.

Radioactive labeling and autoradiography The mouse and hybrid lines were labeled with ³H-thymidine (specific activity 5.9 Ci/mM, Amersham) at a final activity of 0.5 μC/ml. In experiments with the thymidine kinase deficient BI₅₀ cells, ³H-deoxycytidine was substituted as described previously (Graves, 1972b). Chromosome preparations and autoradiographs were prepared as before (Graves, 1972a). Where necessary, the identity of heavily labeled chromosomes could be checked by removing the film in hot water and restaining the slide.

RESULTS

Karyotypes of parent and hybrid lines Hamster chromosomes of groups I and II can all be identified by their relative sizes and arm ratios (as well as some members of groups III and IV), and all the chromosomes of the complement can readily be distinguished from those of either mouse parent (Figure 1). In the mouse complements, only the extra-long and ring markers are identifiable.

The hybrid BN51 between BI₅₀ and NCTC (Graves, 1972b) has almost a full complement of parent chromosomes, and little loss occurred on

subsequent propagation. However, the series of BNa hybrids between BI₅₀ and NCTC(az^F), which evidently underwent a chromosome doubling, showed preferential and progressive loss of hamster chromosomes, and it was possible to isolate clones and subclones retaining virtually all the mouse chromosomes, but with various combinations of from 8 to 20 hamster chromosomes. The pattern of loss was non-random; certain chromosomes (in particular, 1,6 and 9) were retained in all, or nearly all hybrids, while others (particularly of group III) were lost rapidly from most clones.

Of 42 BNa hybrids examined, four were chosen which had retained different sets of hamster chromosomes. All the chromosomes, except 6, 17 and possibly 11 and 18, were absent from at least one of the clones, so it was possible to examine the effect of loss of most of the complement on the rate and sequence of DNA synthesis in remaining hamster chromosomes.

S periods of parent and hybrid lines Table 1 shows average values for the S periods determined from percent labeled mitosis curves as described previously (Graves, 1972b). Both mouse lines have much longer S periods than the hamster parent, and hybrids have S periods similar to those of the mouse parents.

TABLE 1
S periods (hours) of parent and hybrid cell lines

	Parent lines		Hybrids	
	Hamster	Mouse		
BI ₅₀	5.9 ± 0.5	NCTC	8.9 ± 0.4	BN51 9.1 ± 1.1
		NCTC (az ^F)	8.7 ± 0.8	BNa11 8.5 ± 1.5
				BNa57 9.2 ± 1.0
				BNa72 8.8 ± 1.2

Initial DNA synthesis in hybrid cells The grain distribution was studied in cells continuously labeled during the first part of the S period (Graves, 1972b). No striking asynchrony was observed between mouse and hamster complements, even in very lightly labeled cells, which had synthesized labeled DNA probably for only the first few minutes of the S period.

Table 2 shows average grain densities over mouse and hamster complements, calculated from the grain totals, and the summed lengths of mouse and hamster chromosomes in each clone, relative to the length of the extra-long mouse marker. The observed grain totals

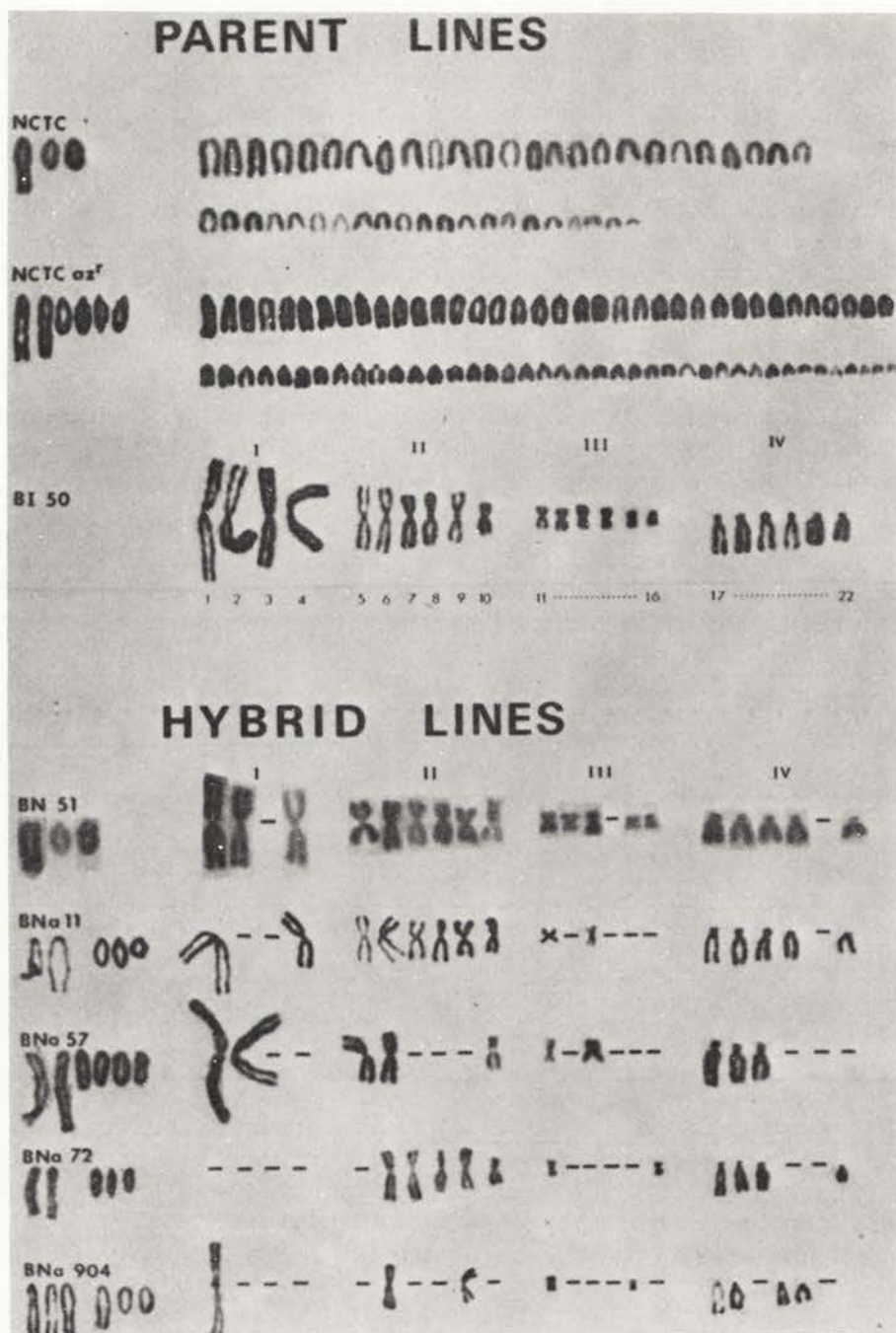


Figure 1. Karyotypes of parent and hybrid cell lines

were tested by χ^2 for significant deviation from the values expected from the ratio of chromosome lengths. In no clone was the deviation significant.

It is concluded that mouse and hamster chromosomes begin DNA synthesis at the same time in all clones.

TABLE 2

Average grain densities over mouse and hamster chromosomes in autoradiographs of cells labeled at the beginning of S phase

Clone	Number of cells scored	Grain density		χ^2_1 on grain totals	Pro-bability
		Mouse	Hamster		
BN 57	10	2.4	2.2	1.384	.3>P>.2
BNa 11	10	1.9	2.1	0.794	.5>P>.3
BNa 57	10	1.4	1.4	0.185	.7>P>.5
BNa 72	10	2.5	2.1	3.159	.1>P>.05
BNa 94	10	1.6	1.8	0.986	.5>P>.3

Terminal DNA synthesis in hybrid cells Cells were labeled continuously for 6 or 8 hours (with colcemid present during the final 4 hours), and were then harvested and prepared for autoradiography. Unlabeled mitoses therefore represented cells which had already completed DNA synthesis before addition of label, lightly labeled cells were just finishing at this time, and progressively more heavily labeled figures synthesized DNA for correspondingly longer periods in the presence of isotope. The order of termination could therefore be deduced.

Asynchrony between mouse and hamster complements

Striking asynchrony was observed between mouse and hamster chromosome sets in all hybrids. The most lightly labeled cells showed grains only over mouse chromosomes (Figure 2a), while in more heavily labeled cells, one or more hamster chromosomes were often unlabeled or lightly labeled (Figure 2b). Thus, hamster chromosomes are first to finish synthesis and mouse chromosomes are last.

Average grain densities over mouse and hamster chromosome sets were calculated as before in lightly labeled cells, and a χ^2 test was again performed on the grain totals. Table 3 shows that in all hybrids, the grain density over mouse chromosomes was higher than over hamster chromosomes. The deviation of grain totals from the expected ratio was highly significant for each clone.

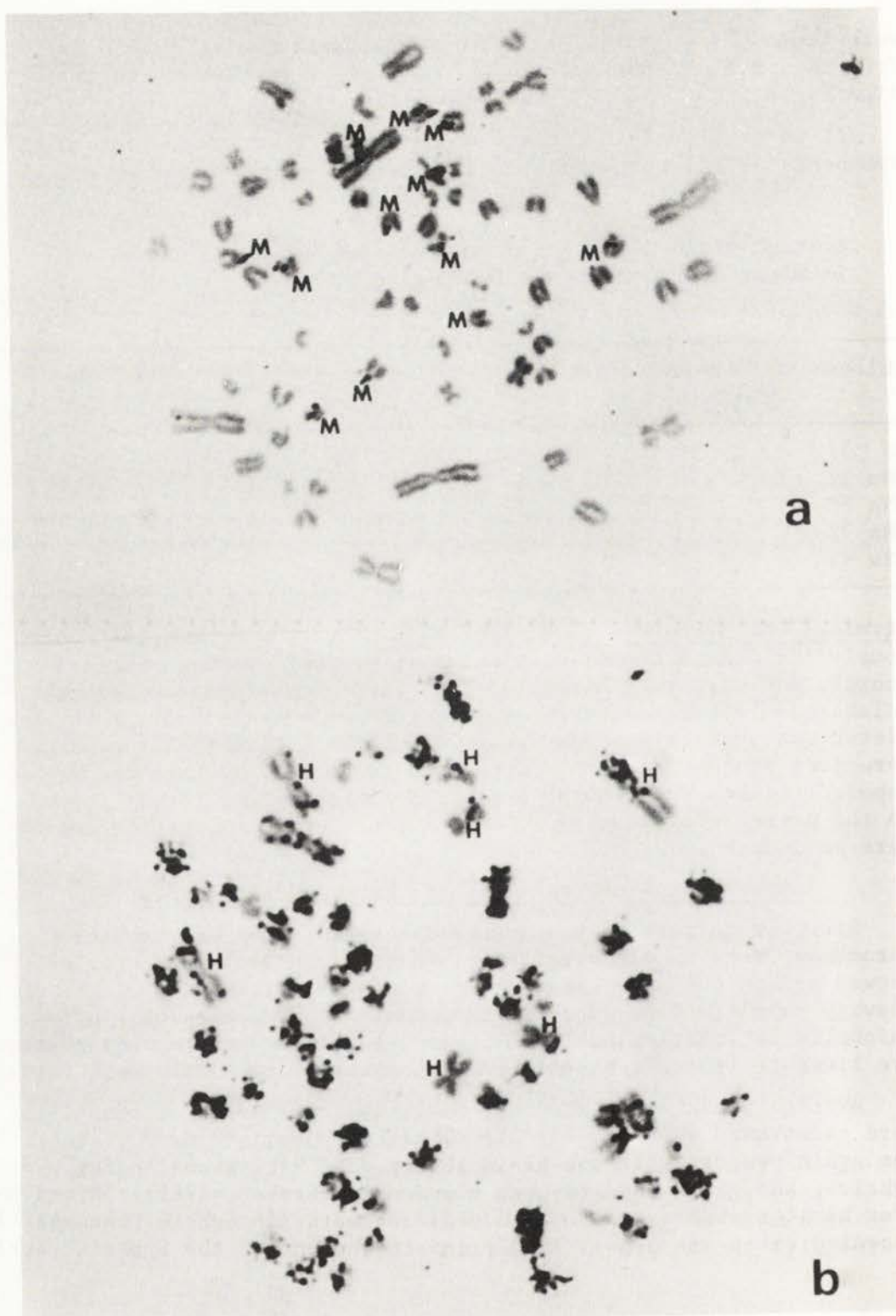


Figure 2. Terminal asynchrony between mouse and hamster chromosome complements

- a. Lightly labeled cell showing late synthesis in 12 mouse chromosomes (M)
- b. Heavily labeled cell showing early termination in hamster chromosomes (H)

It is concluded from these results that hamster chromosomes terminate DNA synthesis considerably earlier than mouse chromosomes, regardless of the chromosome constitution of the cell. There is no evidence, therefore, that loss of any particular hamster chromosome affects the duration of DNA synthesis in the remainder of the complement.

TABLE 3

Average grain densities over mouse and hamster chromosomes in autoradiographs of terminally labeled cells

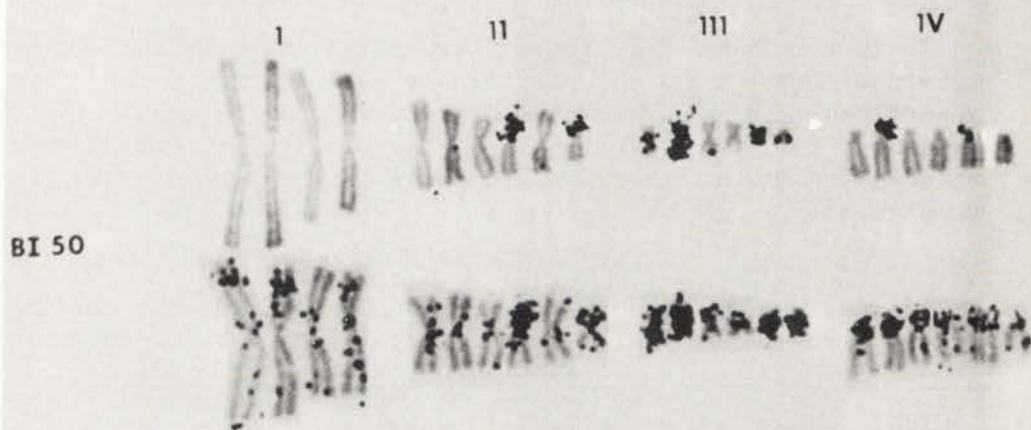
Clone	Number of cells scored	Grain density		χ^2_1 on grain totals	Probability
		Mouse	Hamster		
BN 51	48	2.3	0.7	776.6	$p < .001$
BNa 11	35	2.9	0.8	543.1	$p < .001$
BNa 57	51	2.0	0.5	373.4	$p < .001$
BNa 72	66	1.9	0.6	478.9	$p < .001$
BNa 94	10	2.4	1.4	23.2	$p < .001$

Replication sequence of hamster chromosomes in hamster and hybrid cells

The distribution of grains over hamster chromosomes was compared in autoradiographs of terminally labeled BI₅₀ and hybrid cells (Figure 3). The sequence of termination was found to be very similar, with chromosome 12 and 18 last (where these chromosomes were present) and the short arm of 8, three other group III, and the smaller group IV chromosomes also terminating late. First to complete synthesis were chromosomes 5, 6, 7 and 10 in each clone (where present). The sequence of synthesis within chromosomes was also indistinguishable; the centromere regions of group II, and especially group IV chromosomes terminated much later than the arms, and chromosomes 1 and 2 showed a characteristic late terminating region in the short arm. The short arms of chromosomes 8 and 10 always terminated much later than the long arms.

The course of DNA synthesis in each chromosome during the final hours of the S phase may be represented by histograms, using a differentiation procedure developed previously (Graves, 1967). Since the total number of grains over a cell reflects (though not necessarily linearly) the length of time prior to the end of the S period during

HAMSTER



HYBRIDS

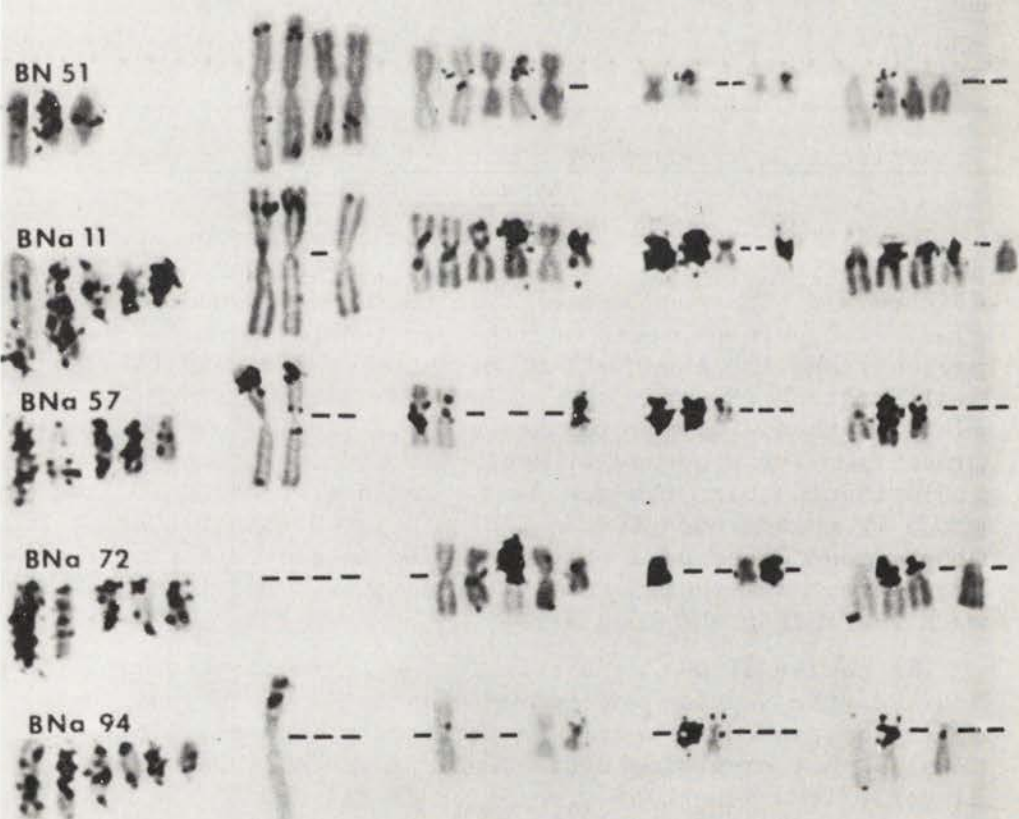


Figure 3. Terminal DNA synthesis sequences of hamster chromosomes in hamster and hybrid cells

which label was available, cells grouped according to their grain totals have accumulated labeled DNA over the same intervals. Cells having a grain total of 1-50 have synthesized labeled DNA during the last interval of the S period (interval 1), while cells having 51-100 grains have accumulated label over intervals 1 and 2, cells with 101-150 grains over intervals 1,2 and 3 etc. To obtain a measure of the amount of DNA synthesized by a particular chromosome during these intervals, the average grain number over this chromosome was calculated for each group (representing cumulative synthesis to the end of S), and differences between successive means were taken (representing the relative amounts of DNA synthesized during each interval). These differences were plotted as histograms for chromosomes 1,6 and 8 in hamster and hybrid cells (Figure 4). Since the hamster complement was not standard for the hybrids, cells in each clone were grouped according to the grain totals over mouse chromosomes, allowing for the difference in chromosome number by using 50 grain increments for the BN51 hybrid and 75 grain increments for the BNa hybrids.

These three chromosomes appeared to follow quite different synthesis patterns in the parent cell. Chromosome 1 showed a decrease in synthesis toward the end of S, while 8 had a late burst of replication and 6 synthesized little DNA during this period. The patterns of synthesis followed by the same chromosomes in hybrid cells were very similar, but were all displaced by 2 or 3 synthesis intervals (as might be expected, since the hamster complement terminated well before the mouse).

The results are consistent with the hypothesis that hamster chromosomes in all hybrids have retained their original sequence of synthesis. Thus there is no indication that loss of a particular chromosome affects the replication pattern in the remainder of the complement.

DNA synthesis in late passage hybrids Several subclones isolated from BNa clones after prolonged growth have recently been karyotyped and autoradiographed. Some of these had greatly reduced hamster complements, others still retained 10 or 12 hamster chromosomes. Preliminary observations of the grain distribution over terminally labeled cells suggests that mouse and hamster sets no longer terminate asynchronously. Since this seems to be the case for hybrids retaining many hamster chromosomes, as well as for those retaining few, it seems unlikely that this change in synthesis pattern can be attributed to loss of particular chromosomes.

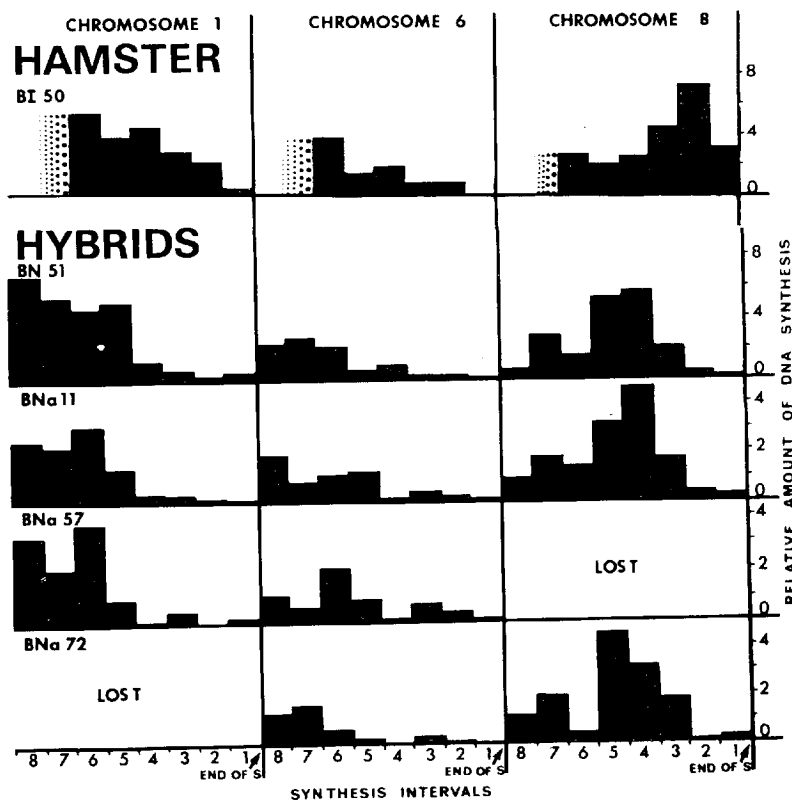


Figure 4. Relative amounts of DNA synthesized by hamster chromosomes 1, 6 and 8 during the last intervals of the S period in hamster and hybrid cells

DISCUSSION

Previous studies with fused cells have shown that the transition from G_1 into the S phase may be induced by a cytoplasmic factor (Rao and Johnson, 1970; Graves, 1972a), but that the duration and sequence of chromosome replication in two unlike nuclei are unaltered by their association in the same cytoplasm (Graves, 1972a). Similar conclusions may be drawn from nuclear transplantation studies with *Amoeba* (Ord, 1969); while S phase cytoplasm contains a factor which initiates DNA synthesis in G_1 phase nuclei, DNA synthesis in S phase nuclei is not inhibited by transplantation into G_1 or G_2 cytoplasm, nor prolonged by transplantation into cells in an earlier stage of S. It appears, therefore, that cytoplasmic factors are involved in initiating DNA synthesis at the beginning of the S period, but play no part in its continuation.

Further work demonstrated also that unlike chromosome sets initiated DNA synthesis synchronously, but thereafter maintained their

original schedule of DNA synthesis, even when combined within the same nucleus of cell hybrids (Graves, 1972b). It was concluded that the rate and pattern of synthesis in the two chromosome sets were controlled autonomously.

Results of the present experiments suggest that this replicative autonomy is a property of individual chromosomes, or smaller units within them. It was observed that the duration and sequence of replication in the chromosomes of reduced mouse x hamster hybrids was not altered by loss of any particular chromosome(s) from the complement (with 2 to 4 possible exceptions, which could not be tested because of their presence in all hybrids).

These studies, therefore, provide no evidence for cytoplasmic, nuclear or chromosomally located factors controlling the pattern of chromosome replication in somatic mammalian cells. Rather, it appears that individual chromosomes contain all the information necessary for the correct timing of their replication.

Other evidence for replicative autonomy of chromosome regions is provided by the observation that the DNA synthesis pattern of a region in the *Drosophila* polytene chromosome is unchanged by translocation (Barr et al., 1968). This demonstrates that DNA synthesis may proceed independently in neighbouring regions, but does not eliminate the possibility that factors located elsewhere in the complement may play some part in control of synthesis in that region.

Although the consistency of the DNA synthesis pattern in chromosome sets of fused cells and hybrids does not indicate that a general cytoplasmic or nuclear control mechanism operates, changes in rate and pattern of DNA synthesis which occur during development may involve just such factors. The S period in *Xenopus* embryos, for instance, lengthens from 15 min in the cleaving egg to several hours by the tail bud stage (Graham and Morgan, 1966), and premeiotic DNA synthesis in *Triturus* was also observed to be enormously prolonged (Callan, 1972). The pattern of DNA synthesis may also change throughout the complement during early development of the mammalian embryo (Hill and Yunis, 1967). Certainly this is true of the inactive X chromosome (Lyon, 1960) in cells from female mammals; during cleavage cycles, the X chromosomes replicate synchronously, but in adult somatic cells, one member of the pair replicates later and faster than the rest of the complement (Comings, 1967; Graves, 1967).

Asynchrony of chromosome replication probably reflects asynchronous DNA synthesis at the molecular level. DNA synthesis in mammalian cells proceeds simultaneously at a number of sites arranged in tandem along a much longer DNA fibre (Huberman and Riggs, 1968). Synthesis is initiated in different replicating units at different

times throughout the S phase (McFarlane and Callan, 1973). Control of the duration and sequence of chromosome replication reflects the control of the sequence of initiation in these units (Callan, 1972).

The mechanism by which initiation in replicons is controlled is quite unclear. There is some evidence (Weiss, 1969) that protein synthesis may be required throughout S, and it is possible that the ordered pattern of initiation could depend on the orderly sequence of synthesis and action of initiator molecules which act on specific replicons (Watson, 1971). The results of the present experiments are quite consistent with such a scheme, providing that the genetic information for these initiator molecules is encoded in the same chromosome as the replicons on which they act. Alternatively, control of initiation may operate by means of structural changes within the chromosome, involving, perhaps, associations with the nuclear membrane (O'Brien et al., 1972). Although studies with hybrid cells provide no evidence for an over-riding control by the nuclear membrane, they do not eliminate the possibility that the nuclear membrane plays a part in ordering the initiation of specific replicating units.

The preliminary observations on late passage hybrids indicate that after prolonged growth, mouse and hamster chromosomes no longer terminate asynchronously. This implies a modification in the duration of the S period in one or other, or both chromosome sets, but until cell cycle studies are performed, it is not known which. Since the change was observed not only in the clones with very few surviving hamster chromosomes, but also in subclones whose karyotypes had undergone little further change, it seems unlikely that it is the result of loss of control factors on particular chromosomes. Rather, it suggests that the overall duration of the S period is controlled by genetic factors, perhaps very many of them, which may be subject to selection toward greater synchrony over a long period of growth.

ACKNOWLEDGMENT

I wish to acknowledge the expert technical assistance of Mrs Clare Greer.

REFERENCES

- Barr, H.J., J.I. Valencia and W. Plaut. 1968. On temporal autonomy of DNA replication in a chromosome translocation. *J. Cell Biol.* 39: 8a.
- Callan, H.G. 1972. Replication of DNA in the chromosomes of eukaryotes. *Proc. Roy. Soc. London B* 181: 19.
- Comings, D.E. 1967. The duration of replication of the inactive X chromosome in humans based on the persistence of the heterochromatic sex chromatin body during DNA synthesis. *Cytogenetics* 6: 20.

- Graham, C.F. and W. Morgan. 1966. Changes in the cell cycle during early amphibian development. *Develop. Biol.* 14: 439.
- Graves, J.M. 1967. DNA synthesis in chromosomes of cultured leukocytes from two marsupial species. *Expt. Cell Res.* 46: 37.
- _____. 1972a. DNA synthesis in heterokaryons formed by fusion of mammalian cells from different species. *Expt. Cell Res.* 72: 393.
- _____. 1972b. Cell cycles and chromosome replication patterns in interspecific somatic hybrids. *Expt. Cell Res.* 73: 81.
- Hill, R.N. and J.J. Yunis. 1967. Mammalian X chromosomes: change in patterns of DNA replication during embryogenesis. *Science* 155: 11.
- Huberman, J.A. and A.D. Riggs. 1968. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* 32: 327.
- Littlefield, J.W. 1964. Selection of hybrids from matings of fibroblasts *in vitro* and their presumed recombinants. *Science* 145: 709.
- Lyon, M. 1961. Gene action in the X chromosome of the mouse (*Mus musculus* L.). *Nature* 190: 372.
- Martin, P.G. 1966. The pattern of autosomal DNA replication in four tissues of the Chinese hamster. *Expt. Cell Res.* 45: 85.
- McFarlane, P.W. and H.G. Callan. 1973. DNA replication in the chromosomes of the chicken, *Gallus domesticus*. *J. Cell Sci.* 13: 821.
- Mueller, G.C. and K. Kajiwara. 1966. Early and late-replicating deoxyribonucleic acid complexes in HeLa nuclei. *Biochim. Biophys. Acta* 114: 108.
- O'Brien, R.L., A.B. Sanyal and R.H. Stanton. 1972. Association of DNA replication with the nuclear membrane of HeLa cells. *Expt. Cell Res.* 70: 106.
- Ord, M.J. 1969. Control of DNA synthesis in *Amoeba proteus*. *Nature* 221: 964.
- Rao, P.N. and R.T. Johnson. 1970. Mammalian cell fusion: Studies on the regulation of DNA synthesis and mitosis. *Nature* 225: 159.
- Watson, J.D. 1971. The regulation of DNA synthesis in Eukaryotes. *Adv. Cell Biol.* 2: 1.
- Weiss, B.G. 1969. The dependence of DNA synthesis on protein synthesis in HeLa S3 cells. *J. Cell Physiol.* 73: 85.

Sex chromosome evolution and activity in mammals, particularly kangaroos

D. W. Cooper, P. G. Johnston, C. E. Murtagh and J. L. VandeBerg

It is well established that in eutherian mammals dosage compensation for X chromosomes in the female is achieved by a mechanism called random X-inactivation or Lyonization. In the soma, only one X is active in any one cell. Some cells have an active maternal X, others an active paternal X, the two types of cell generally being in an approximately 1:1 ratio. Marsupials, or kangaroos at least, are different. In several species of kangaroos and wallabies three kinds of X chromosome marker are available: (a) X chromosomes of different sizes; (b) electrophoretic variants of glucose-6-phosphate dehydrogenase (G6PD); (c) electrophoretic variants of phosphoglycerate kinase (PGK). The study of these markers has yielded these results: (1) The X chromosome of kangaroos is at least partially homologous with that of eutherians. (2) In lymphocytes and the nucleated precursors of erythrocytes there is only one active X, as in eutherians, but unlike eutherians, it is always the maternal X. The paternal X is not manifested in any detectable degree. These cells have paternal X-inactivation rather than random X-inactivation. (3) Work on PGK in muscle tissue and mass cultures of fibroblasts shows that in these cells both X chromosomes are active. However, the maternal chromosome is always more strongly expressed than the paternal one. There appears to be partial paternal X-inactivation. Our current work is directed towards discovering whether this last result means that both X chromosomes are active in the one cell in unequal degree, or whether muscle and populations of cultured fibroblasts are a mosaic of two kinds of cells, the majority with active maternal chromosomes and the minority with active paternal chromosomes.

Since 1961, when Lyon put forward her hypothesis of random X-inactivation as the basis of sex chromosome dosage compensation in eutherian mammals, a great deal of interest has been taken in the mammalian X chromosome as an example of a coordinately controlled unit of genetic activity. Lyon has discussed information and ideas about X-inactivation in several reviews, the most recent being Lyon (1972). The principal features of the system in eutherian mammals are (a) inactivation at random of either the paternal or maternal X at a very early stage of development, possibly the late blastocyst stage; (b) completeness and irreversibility of inactivation; (c) late DNA synthesis of the inactive X; (d) the existence of the inactive X during interphase as the condensed chromatin body, the Barr body or sex chromatin; and (e) the fact that only

one X in humans with two or more X chromosomes is active. Since marsupials have fewer chromosomes than eutherian mammals and are generally favourable for cytogenetic work (Sharman, 1961, 1973; Hayman and Martin, 1969, 1974), their system of X chromosome dosage compensation has been investigated. The data obtained so far support Ohno's (1967, 1969) hypothesis that the X of mammals has been a conservative linkage group in mammalian evolution. They also show that kangaroos have preferential paternal X-inactivation rather than random X-inactivation. The latter conclusion was drawn from the results of studying the timing of DNA synthesis in the lymphocytes of hybrid kangaroos with X chromosomes of different sizes (Sharman, 1971) and from studies on sex linked enzymes (Richardson et al., 1971; Cooper et al., 1971). Some recent data show that in muscle and primary uncloned cultured cells from females, both X linked genes may be expressed (VandeBerg et al., 1973b; Cooper et al., 1974).

THE EXPRESSION OF X-LINKED MARKERS IN KANGAROOS

The information on the electrophoretic phenotypes and inheritance of two sex linked enzyme markers, G6PD and PGK A, in kangaroos has been reviewed in detail (Cooper et al., 1974). An outline of the principal results is given here.

One of the two electrophoretically separable allozymic forms of G6PD, designated F and S, may be found in the blood and tissues of kangaroos of the genus *Macropus* listed in Table 1. These two types are depicted in Figure 1. All kangaroos have only one of these two types in their erythrocytes. No kangaroo has been found with more than one, even in trace amounts. If G6PD were inherited as a sex linked trait as in man, one would expect to find heterozygous females with both allozymes. However, both females and males which are the result of crosses between species or between subspecies monomorphic for contrasting allozymes show one allozyme only, that contributed by their mother. For euros and wallaroos, this cross has been made reciprocally. These data are compatible with the hypothesis of sex linkage of the gene with inactivation of the paternal X (Richardson et al., 1971; Sharman, 1973; Sharman and Johnston, 1973). The paternal gene which is not expressed in the blood may be transmitted to the next generation and there expressed. Thus inactivation may not occur in the germ line, or if it does, reactivation must take place. A true polymorphism, i.e. variation within a population, exists for G6PD in the red-necked wallaby (Johnston, VandeBerg and Sharman, unpublished). Only F or S individuals are found in both sexes. Limited breeding data are consistent with sex linkage and paternal X-inactivation, although it is not possible to rule out autosomal inheritance with one of the two forms being dominant.

TABLE 1

Species, hybrids, X-linked allozyme markers and X chromosomes of different sizes used to study X-inactivation in kangaroos

-
- 1 *Macropus giganteus* Shaw, 1790 (Eastern Grey Kangaroo). Polymorphic for PGK A with the variant VE having a frequency of approximately 10%
 - 2 *Macropus parryi* Bennett, 1835 (Pretty-face or Whiptail Wallaby). Polymorphic for PGK A with the variant VP having a frequency of approximately 55%
 - 3 Crosses between *Macropus robustus robustus* Gould, 1841 (Wallaroo) and *Macropus robustus erubescens* Sclater, 1870 (Euro). The wallaroo is monomorphic for the F G6PD type and has a small X; the euro has G6PD S and a larger X. The F1 is fertile and can be backcrossed to both parental species
 - 4 A species hybrid between a female wallaroo (see 3 above) and *Macropus rufus* Desmarest, 1822 (= *Megaleia rufa*) (Red Kangaroo). The red kangaroo has G6PD S and an X larger than either subspecies of *M. robustus*
 - 5 *Macropus rufogriseus banksianus* Quoy and Gaimard, 1825 (Red-necked wallaby). Polymorphic for G6PD, having both F and S in some populations
-

The results for the PGK A polymorphism parallel those for G6PD. Sex linked electrophoretic variation for PGK A exists in man; females heterozygous express both allozymes in their blood (Chen et al., 1971; Chen and Giblett, 1972). But in kangaroos one only of several allozymes may be found in the erythrocytes of a particular individual (Cooper et al., 1971; VandeBerg et al., 1973b). In the eastern grey kangaroo there are two, N and VE; differences in the ratio of N:VE were found in the progeny from reciprocal crosses between N and VE. This is suggestive of X linkage with paternal X-inactivation but does not constitute proof. However, the population data for muscle phenotypes of the pretty-face wallaby do establish sex linkage for this enzyme (see below). As with G6PD, the PGK A allele which is not expressed in the blood may be transmitted to the next generation in active form.

Sharman and Johnston (1973) and Johnston (unpublished) have examined the tissue homogenates of four female pouch young which were heterozygous for G6PD. The animals were crosses and backcrosses

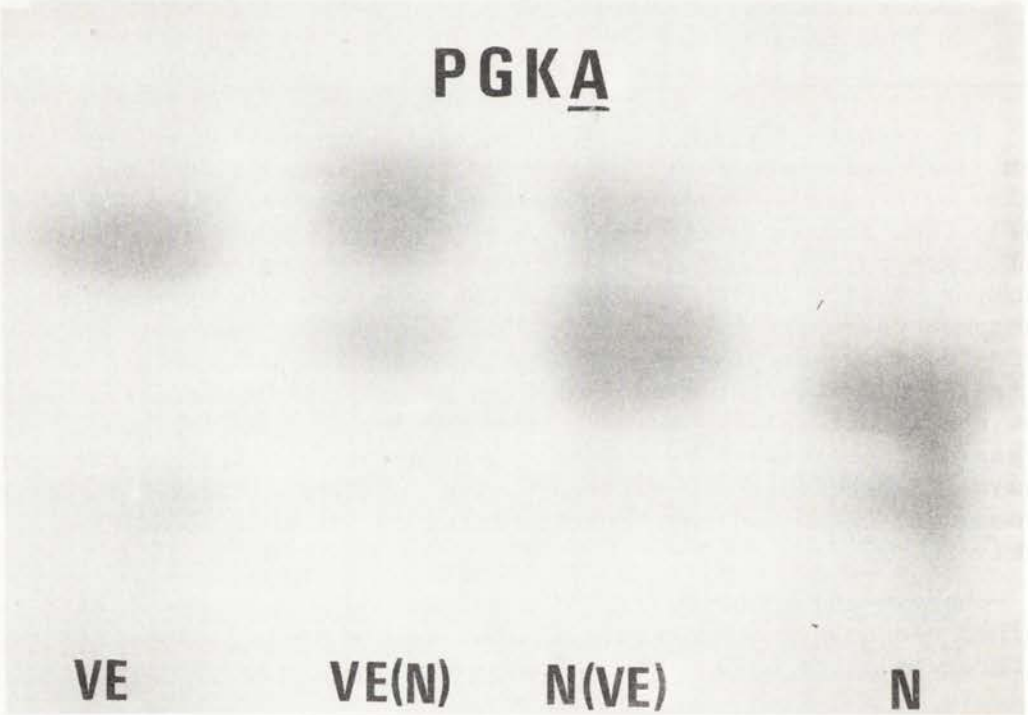
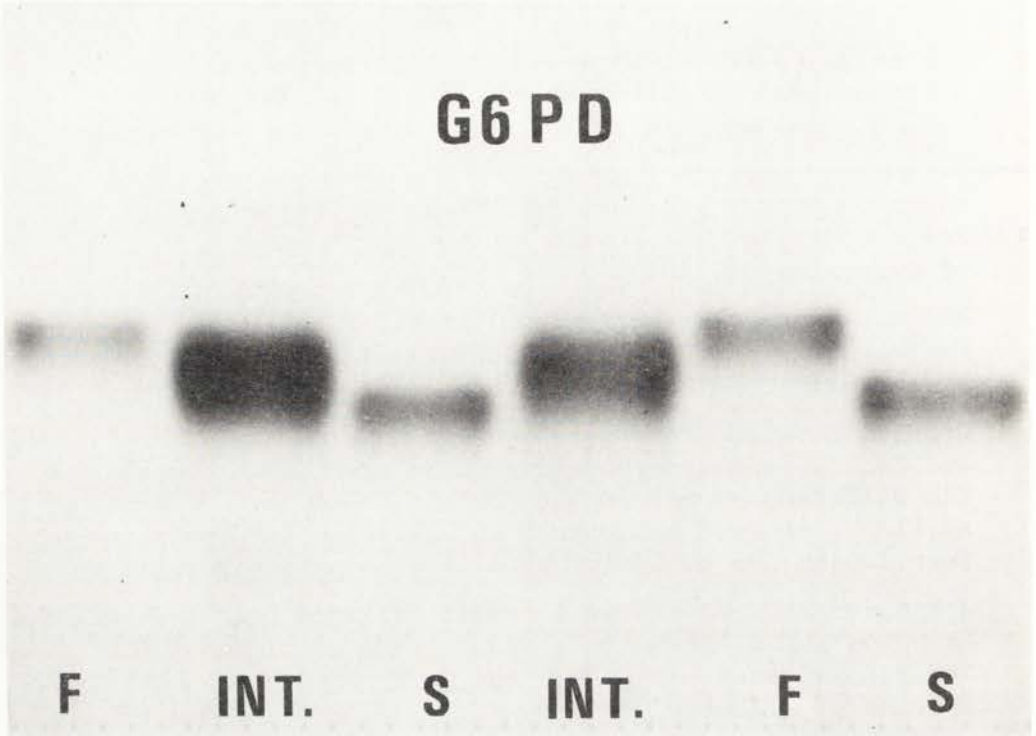


Figure 1. Top. G6PD variation in kangaroos. F (Fast) is found in *Macropus robustus robustus* and in most individuals of *Macropus rufogriseus banksianus*. S (Slow) is found in a minority of *banksianus* individuals and all the other animals mentioned in the text. Int (Intermediate) is found in cultured fibroblasts from females heterozygous for G6PD variation i.e., hybrids produced by crosses between euros and wallaroos, some backcrosses and in a wallaroo red kangaroo hybrid. The intermediate type is probably the result of the formation of interaction products.

Bottom. PGK A electrophoretic variation in the eastern grey kangaroo. VE is a variant found in 10% of individuals of this species, N is the normal type found in most animals of the *Macropus* genus. VE (N) is the phenotype seen in cultured cells of an animal produced by a VE (♀) x N (♂) cross while N (VE) is the phenotype from the reciprocal cross. Note how the maternally contributed form is predominant. The faint staining slower migrating form is PGK B, a minor PGK component of the blood which is autosomally inherited. PGK B is absent from the fibroblasts and is the sole enzyme of sperm (VandeBerg et al., 1973a).

between euros and wallaroos. Only the maternally contributed allele was expressed in all the tissues examined, namely ovaries, uteri, kidney, liver, heart and skeletal muscle, as well as in blood. Complete paternal X-inactivation appears to operate for G6PD in *Macropus robustus* at the pouch young stage.

The results which have been obtained for PGK A in tissue homogenates of the pretty-face wallaby *Macropus parryi* are somewhat different (VandeBerg et al., 1973b). This species has two allozymes N and VP (Table 1). All twenty males examined had the same allozyme throughout all their tissues, with no hint of the appearance of the other form. Twelve of 24 adult females were the same in this respect. The other 12 adult females had a number of tissues in which one of the two allozymes was manifested to the usual extent with trace amounts of the other form. Six had full expression of N with trace VP, and six full expression of VP with trace N. Cardiac and skeletal muscle and one smooth muscle, the bladder, had these phenotypes while liver, spleen, kidney, intestine, vaginae and ovaries and uteri had only the one type. Seven of the 12 females which had both allozymes manifested were known to be heterozygous because they carried pouch young with the allozyme unlike their own, i.e. the female had transmitted her presumably paternally derived allele to her offspring in active form. These results are compatible with sex linkage and paternal X-inactivation, although there is no information on parents of the females. They are not easily explained by other hypotheses. We have also examined the PGK A allozymes of a 35-day old eastern grey kangaroo female heterozygous at this locus. This had expression of its paternally derived allele in the reproductive tract and cardiac muscle but not the skeletal muscle (Cooper et al., 1971). The difference between this result and the pretty-face data may be due either to a developmental or a species difference. We did

not find any pretty-face pouch young females which had expression of the presumed paternal allele, but their level of PGK A was so low that the trace amounts found in adults would probably have gone undetected.

Fibroblasts have been cultured from ear explants and leg muscle fascia for females heterozygous for either G6PD or PGK A, along with appropriate homozygous or hemizygous controls. For both G6PD and PGK A there appears to be expression of the paternally as well as the maternally derived allele. The PGK A results parallel those found in the muscle of the pretty-face wallaby (Figure 1). Two eastern grey kangaroo heterozygotes, one from an N (♀) x VE (♂), the other from the reciprocal cross, had both forms manifested in a culture of ear origin, but with the maternally contributed form predominant. A culture set up from the ear of a pretty-face female of unknown parents had VP with a trace of N. Since PGK A is a monomer, interaction products are not possible and one cannot say if both allozymes are being produced in the same cell or if these cultures are a mosaic of cells producing contrasting allozymes.

The G6PD patterns in cultured cells of heterozygotes have neither the F nor S pattern seen in all other cells, but instead a form we have given the provisional designation of Intermediate (Figure 1). G6PD may be either a dimer or tetramer (see Cooper et al., 1974 for references) and our interpretation of this pattern is that interaction products are being formed. Mixing experiments indicate that the possibility that these are being generated *in vitro* when the cells are being prepared for electrophoresis is a remote one. This leaves three possibilities. The cultures may consist of a mixture of cells, some with only an allele for F active, others with only the allele for S active, with the cells exchanging either G6PD monomer or mRNA to allow the formation of the interaction products. This is not likely. We have grown *Macropus rufogriseus* (G6PD F) and *Macropus giganteus* (G6PD S) together in mixed culture for four passages and the intermediate form was never found. Tetraploid cells do occur in these cultures in frequencies up to 25% in some high passages. If such cells arise by fusion between diploid cells in a mosaic, some interaction product will be made. As a general explanation of the results this is ruled out by the fact that cultures which are largely diploid show the Intermediate pattern. Or finally, both alleles may be being expressed in the one cell. This third possibility seems rather more likely, but until we can clone these cells, no definite answer can be given for either enzyme. The G6PD patterns do not have the asymmetry characteristic of the PGK A patterns, suggesting that both paternally and maternally derived G6PD alleles may be equally manifested in the cultured cells.

Table 2 summarizes the principal results which have been obtained to date on the activity of the X linked allozymes in various kangaroo cells and tissues. It is clear that complete paternal X-inactivation is not characteristic of all somatic cells. The results for PGK A and G6PD are not completely concordant. Since the variation for the two enzymes occurs in different species, one cannot say if we are dealing with a difference between loci or a difference between species. Moreover, most of the G6PD tissue data concern pouch young while PGK A tissue data are for adults. More sex linked markers and a species with at least two polymorphic loci are needed to resolve these questions.

TABLE 2

Expression of X chromosomes in some tissues and cells of kangaroos

Cell type	Activity of		Evidence
	X ^m	X ^p	
Female germ line	+	+ at some point	Paternally derived X-linked allozyme markers and marker X chromosomes may be transmitted to the next generation in active form
Male germ line	+ ¹		X becomes heterochromatic at meiosis I
Blood	+	-	PGK A, G6PD, DNA synthesis in several species
Kidney, liver	+	-	PGK A, in <i>M. parryi</i> , <i>M. giganteus</i> and G6PD in <i>M. robustus</i>
Skeletal and cardiac muscle			
(1) PGK A	+	+ ²	PGK A in <i>M. parryi</i>
(2) G6PD	+	-	G6PD in <i>M. robustus</i>
Cultured fibroblasts (ear, fascia)	+	+ ³	PGK A, <i>M. giganteus</i> , <i>M. parryi</i> and G6PD in <i>M. robustus</i> / <i>M. rufus</i> hybrids and in <i>M. robustus</i>

¹Until first division of meiosis (see McIntosh and Sharman, 1953).

²Partial expression only.

³Partial expression for PGK A, possibly full expression for G6PD.

The G6PD results for cultured fibroblasts suggest that both X chromosomes may be active within the one cell. That cells cultured from eutherians have only one active X is well established (Lyon, 1972). Cells with two active X chromosomes will be difficult to accommodate in any of the models of X-inactivation which have so far been suggested (Cooper, 1971; Lyon, 1971, 1972; Brown and Chandra, 1973).

THE CONSERVATION OF THE MAMMALIAN X-LINKAGE GROUP : OHNO'S HYPOTHESIS

In 1967, Ohno advanced the hypothesis that the X chromosomes of mammals all contain the same set of homologous genes. He suggested that a gene found to be sex linked in one species of mammal could be assumed to be sex linked in all others. He was led to formulate this hypothesis because he had shown that one X constitutes a constant proportion, approximately 5%, of the total haploid genome in a wide variety of mammalian species (Ohno et al., 1964). Since 1967, a great deal of information has been collected which supports Ohno's hypothesis, and only two possible exceptions have been discovered. In this section we shall summarize the two main lines of evidence in its favour.

X-autosome translocations The existence of the mammalian dosage compensation mechanism is itself a reason for believing that Ohno's thesis is at least plausible, because any translocation of autosomal genes to the X, or X-linked genes to an autosome, could lead to an upset in the dose relationships for the translocated segments. The behaviour of X-autosome reciprocal translocations suggests that Ohno's hypothesis may be correct. Such translocations lead to infertility of males carrying them in the mouse (Eicher, 1970; Russell and Montgomery, 1969, 1970) and to a reduction in fertility in man (Buckton et al., 1971). The only X-autosome translocation in the mouse which is compatible with fertility is Cattanaich's (1961) *flecked* translocation (T(1 : X) Ct). This is an insertion of a piece of autosome into the X, without any translocation of X material to the autosome. If infertility as a result of X-autosome reciprocal translocations is a general phenomenon in male mammals, it would be difficult for such rearrangements to become established in all members of a species. Ohno (1969) states that the only X-autosome translocations which have been so established are whole X-whole autosome translocations, which lead to an XY_1Y_2/XX male/female dimorphism. The mammals in which such dimorphisms occur are listed in Fredga (1970). In the two marsupial species in which this has occurred (*Wallabia bicolor* (= *Protomnodon bicolor*) and *Potorous tridactylus*, Sharman, 1961), only the original X in the composite chromosome may be late labelling in the female (Hayman and Martin, 1965), i.e. the dosage compensation mechanism is preserved.

What system causes X-autosome reciprocal translocations to be male sterile? In a recent paper, Lifschytz and Lindsley (1972) suggest a possibility. They postulate that the X chromosome in all male heterogametic organisms is normally inactivated during a critical stage in spermatogenesis. An X-autosome reciprocal translocation causes sterility because it leaves at least one part of the X active. The behaviour of the X in reciprocal translocations during male meiosis postulated by Lifschytz and Lindsley is very similar to the behaviour of such translocations in female mice. Russell (1963) and Russell and Montgomery (1969) have proposed that the two parts of the X separated by translocation behave differently. One part remains susceptible to inactivation and inactivates autosomal genes adjacent to the breakpoint on the attached autosomal segment; the other part is neither inactivated nor inactivates any genes on its attached autosomal material. Russell postulates an inactivation centre on the part of the X which controls both the ability to inactivate and the capacity to be inactivated. Inactivation is said to proceed outward in a gradient from this centre, so that the extremities of the chromosome may be partially or wholly active. Lifschytz and Lindsley speculate that the inactivation which occurs in primary spermatocytes and that which occurs in the somatic cells of female mammals are brought about by the same or similar mechanisms. They observe that the fact that kangaroos possess paternal X-inactivation suggests a connection between the two phenomena.

Despite the fact that both the occurrence of sex-linked traits in a wide variety of mammalian species and the male sterility of X-autosome reciprocal translocations lend strong support to Ohno's hypothesis, there are reasons for testing it further for certain species. Paradoxically, one reason is the converse of that which led Ohno to put forward his hypothesis; there are some marsupial and eutherian species which have an X chromosome which differs markedly in size from the 'standard' 5% of the haploid genome. Ohno (1969) and Fredga (1970) list the eutherian species with unusually large X chromosomes which are not the product of a translocation of an entire X onto an autosome. Within macropodids (kangaroos and their relatives) the X varies considerably. For example *Macropus eugenii* (= *Protemnodon eugenii*) has 5% of its total arm length per haploid genome in the X, and *Macropus rufus* 10% (Martin and Hayman, 1965). Larger macropodid X chromosomes exist. From the relative length of X chromosomes in Figure 7 of Hayman and Martin (1969), it would appear that the X chromosome of *Macropus rufogriseus* is approximately 12%, and that of *Macropus parryi* approximately 20% of the total haploid genome. The common wombat *Phascolomis mitchelli* has an X of 9% (Martin and Hayman, 1967). Ohno (1969) suggests that such large X chromosomes in eutherians are the result of duplications, triplications and quadruplications

with the extra material so generated becoming constitutively heterochromatic.

There are also species with small chromosomes. Amongst Australian marsupials *Cercatetus nanus* and *Cercatetus concinnus* have an X which constitutes only 2% of the total haploid genome, and all dasyurid X chromosomes are between 3 and 4% (Martin and Hayman, 1967). Do these species have fewer sex-linked genes? The only region for which an answer to this question can be given is the nucleolar organiser which presumably contains the genes for the ribosomal RNA sequences. This is indubitably on the X in all except one of the many kangaroos which have been examined. In none of the non-macropodid marsupials has a nucleolar organiser been found on the X chromosome and several dasyurid, peramelid, caenolestid and burramyid marsupials have secondary constrictions (presumably nucleolar organisers) on an autosomal pair (summarized in Sharman, 1973). The question of whether the nucleolar organiser is subject to paternal X-inactivation in kangaroos has not been answered. In the eutherian *Microtus agrestis*, both the X chromosomes in cultured fibroblasts appear to have active nucleolar organisers (Natarajan and Sharma, 1971). If so, how is dosage compensation achieved for ribosomal RNA genes? Clearly, the male cannot make half as many ribosomes as the female makes. Some dosage compensation mechanism must operate, although it may not be the single active X system which governs the rest of the chromosome.

Homologous sex-linked traits The obvious way of testing the hypothesis is to look for variation in an enzyme, protein or antigen known to be sex linked in one species and see if such traits are sex linked in others. Ohno (1969) was able to list six different traits sex linked in at least two species from six eutherian mammalian orders. Since that time, several new examples of homology have been added for eutherian species (see Lyon, 1972 for references). As outlined in this paper, G6PD and PGK are both sex linked in several eutherian species and one marsupial group, kangaroos of the genus *Macropus*. This is rather striking because the marsupial and eutherian lineages are estimated to have diverged 130 million years ago (Air et al., 1971). G6PD and PGK contain only the one kind of polypeptide, i.e. are controlled by the one structural locus. The variation is almost certainly in their primary structure, since it is electrophoretic. The demonstration that homologous genes in these widely separated species are sex linked is accordingly quite rigorous.

The enzyme α -galactosidase is a possible exception to Ohno's hypothesis. It is absent from individuals with the sex linked Fabry's disease (Kint, 1970). When hamster-human somatic cell hybrids are formed, the hybrids possess the hamster electrophoretic form of the enzyme and another electrophoretic form which is slightly faster

than the human enzyme. The human enzyme itself does not appear in the hybrids. It is possible that the form peculiar to the hybrids is a post-synthetic modification of the human form. In any case, this form disappears when the human X is lost from the hybrid cells. This result suggests but does not conclusively prove that a structural locus for human α -galactosidase is on the X (Grzeschik et al., 1972). Beutler and Kuhl (1972) have shown that male and female mules possess both the donkey and horse electrophoretic forms of the enzyme. They cannot rule out the possibility that the difference between horse and mule is the result of a post-synthetic modification, but it seems most likely that a structural locus for the enzyme is autosomal in the *Equidae*. The resolution of this problem must await a detailed analysis of the structure of the enzyme.

It should be evident from the discussion in this paper that further understanding of the evolution of the mammalian X and of its dosage compensation mechanism could be arrived at by the discovery of new X-linked markers. In view of the fact that two of the three major groups of mammals, marsupials and monotremes, are most easily accessible in Australia, we intend to continue the search for such markers, particularly allozymic markers. We hope that a combination of these markers and the low chromosome number of marsupials may also allow analysis of the dosage compensation mechanism by the techniques of somatic cell genetics.

ACKNOWLEDGMENTS

Some of the work referred to in here was supported by Australian Research Grants Committee grants to G.B. Sharman and D.W. Cooper. We thank Professor G.B. Sharman for help in preparation of the manuscript.

REFERENCES

- Air, G.M., E.O.P. Thompson, B.J. Richardson and G.B. Sharman. 1971. Amino acid sequences of kangaroo myoglobin and haemoglobin and the date of marsupial-eutherian divergence. *Nature* 229: 391.
- Beutler, E. and W. Kuhl. 1972. Biochemical studies of α -galactosidase in man. *Amer. J. Hum. Genet.* 24: 237.
- Brown, S.W. and H.S. Chandra. 1973. Inactivation system of the mammalian X chromosome. *Proc. Nat. Acad. Sci. USA* 70: 195.
- Buckton, K.E., P.A. Jacobs, C.A. Rae, M.S. Newton and R. Sanger. 1971. An X-autosome translocation in man. *Ann. Hum. Genet.* 35: 171.
- Cattanach, B.M. 1961. A chemically-induced variegated-type position effect in the mouse. *Z. VererbLehre* 92: 165.
- Chen, S-H. and E.R. Giblett. 1972. Phosphoglycerate kinase : Additional variants and their geographic distribution. *Amer. J. Hum. Genet.* 24: 229.

- Chen, S-H., L.A. Malcolm, A. Yoshida and E.R. Giblett. 1971. Phosphoglycerate kinase polymorphism in man. *Amer. J. Hum. Genet.* 28: 87.
- Cooper, D.W. 1971. A directed genetic change model for X-inactivation in eutherian mammals. *Nature* 230: 292.
- Cooper, D.W., J.L. Vandenberg, G.B. Sharman and W.E. Poole. 1971. Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. *Nature New Biol.* 230: 155.
- Cooper, D.W., P.G. Johnston, C.E. Murtagh, G.B. Sharman, J.L. Vandenberg and W.E. Poole. 1974. Sex-linked allozymes and sex chromosome evolution and inactivation in kangaroos. *Proc. Third Int. Isozyme Conf.* Yale University. Academic Press. In press.
- Eicher, E.M. 1970. X-autosome translocations in the mouse: Total inactivation versus partial inactivation of the X-chromosome. *Adv. Genet.* 15: 175.
- Fredga, K. 1970. Unusual sex chromosome inheritance in mammals. *Phil. Trans. Roy. Soc. London B* 259: 15.
- Grzeschik, K.H., S. Grzeschik, S. Banhof, G. Romeo, M. Siniscalco, H. van Someren, P. Meera Khan, A. Westerveld and D. Bootsma. 1972. X-linkage of human α -galactosidase. *Nature New Biol.* 240: 48.
- Hayman, D.L. and P.G. Martin. 1965. An autoradiographic study of DNA synthesis in the sex chromosomes of two marsupials with an XX/X₁Y₂ sex chromosome mechanism. *Cytogenetics* 4: 1201.
- _____. 1969. Cytogenetics in marsupials. In *Comparative Mammalian Cytogenetics* (K. Benirschke, ed.) p.191. Springer-Verlag, Berlin.
- _____. 1974. *Animal Cytogenetics. Volume 4. Monotremata and Marsupialia.* Published by Gerbrüder Bornträger. Berlin-Stuttgart. In press.
- Kint, J.A. 1970. Fabry's disease: Alpha-galactosidase deficiency. *Science* 167: 1268.
- Lifschytz, E. and D.L. Lindsley. 1972. The role of X-chromosome inactivation during spermatogenesis. *Proc. Nat. Acad. Sci. USA* 69: 182.
- Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190: 372.
- _____. 1971. Possible mechanisms of X-chromosome inactivation. *Nature New Biol.* 232: 229.
- _____. 1972. X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev.* 47: 1.
- Martin, P.G. and D.L. Hayman. 1965. A quantitative method for comparing the karyotypes of related species. *Evolution* 19: 157.
- _____. 1967. Quantitative comparisons between the karyotypes of Australian marsupials from three different superfamilies. *Chromosoma* 20: 290.
- McIntosh, A.J. and G.B. Sharman. 1953. The chromosomes of some species of marsupials. *J. Morphol.* 93: 509.

- Natarajan, A.T. and R.P. Sharma. 1971. Tritiated uridine induced chromosome aberrations in relation to heterochromatin and nucleolar organisation in *Microtus agrestis* L. *Chromosoma* 34: 168.
- Ohno, S. 1967. Sex chromosomes and sex linked genes. Springer-Verlag, Berlin, Heidelberg and New York.
- _____. 1969. Evolution of sex chromosomes in mammals. *Ann. Rev. Genet.* 3: 495.
- Ohno, S., W. Becak and M.L. Becak. 1964. X-autosome ratio and the behaviour pattern of individual X-chromosomes in placental mammals. *Chromosoma* 15: 14.
- Richardson, B.J., A.B. Czappon and G.B. Sharman. 1971. Inheritance of glucose-6-phosphate dehydrogenase variation in kangaroos. *Nature New Biol.* 230: 154.
- Russell, L.B. 1963. Mammalian X chromosome action : Inactivation limited in spread and in region of origin. *Science* 140: 976.
- Russell, L.B. and C.S. Montgomery. 1969. Comparative studies on X-autosome translocations in the mouse. I. Origin, viability, fertility and weight of five T(X; 1)s. *Genetics* 63: 103.
- _____, _____. 1970. Comparative studies on X-autosome translocations in the mouse. II. Inactivation of autosomal loci, segregation and mapping of autosomal breakpoints in five T(X; 1)s. *Genetics* 64: 281.
- Sharman, G.B. 1961. The mitotic chromosomes of marsupials and their bearing on taxonomy and phylogeny. *Aust. J. Zool.* 9: 38.
- _____. 1971. Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature* 230: 231.
- _____. 1973. The chromosomes of non eutherian mammals. In *Cytotaxonomy and Vertebrate Evolution* (A.B. Chiarelli and E. Capanna, ed.). Academic Press, London and New York.
- Sharman, G.B. and P.G. Johnston. 1973. X-chromosome and X-linked gene inactivation in marsupials. *Genetics* 74: 250.
- VandeBerg, J.L., D.W. Cooper and P.J. Close. 1973a. Mammalian testis phosphoglycerate kinase. *Nature New Biol.* 243: 48.
- VandeBerg, J.L., D.W. Cooper and G.B. Sharman. 1973b. Phosphoglycerate kinase A polymorphism in the wallaby *Macropus parryi* : Activity of both X chromosomes in muscle. *Nature New Biol.* 243: 47.

Heterochromatin and timing of DNA replication in morabine grasshoppers

G. C. Webb and M. J. D. White

Morabine karyotypes include one small autosomal pair (usually free, but fused with other chromosomes in some species) which is positively heteropycnotic in interphase and which replicates later and faster than euchromatin. Morabines show negative heteropycnosis of the X during early spermatogonial divisions, in which they start and finish replication later than the euchromatin. In the somatic cells the X chromosomes are only moderately late-replicating, to varying degrees. In one species two cysts of tetraploid spermatogonia showed asynchrony of the two X chromosomes, only one of which was negatively heteropycnotic. In the parthenogenetic species *Moraba virgo*, in which crossing-over is restricted to the identical products of chromosome doubling, the longest autosomal pair and the X₁ pair are extensively heterozygous for late-replicating segments. These segments do not show allocyly except for the short arm of one of the X₁ chromosomes which forms interphase heterochromatin and shows close chromatid apposition during prophase.

A considerable variety of different types of allocyly and asynchronies of DNA replication has been found in the Australian grasshoppers of the subfamily Morabinae, a group containing approximately 250 species most of which are as yet undescribed. There are several different kinds of heterochromatin in the karyotypes of these insects. The majority of morabine species have $2n \delta = 17$, $2n \text{♀} = 18$ and an XO:XX type of sex determining mechanism. But a large number of chromosomal rearrangements, including at least 39 fusions and 22 dissociations, have occurred in the phylogeny, giving rise to a great variety of karyotypes.

For autoradiography the insects were injected with tritiated thymidine followed by colcemid after a period which allowed for the elapse of the average G2 period of the cells. The cells which were most likely to be dividing in the walls of the ovarian follicles of adult females had an average G2 of 10 hours at 32°C (Figure 1). In mature ovipositing females large groups of cells with a high mitotic index and a relatively short G2 were occasionally present; these afforded an opportunity to observe chromosomes labelled in early and mid S-phase (they were probably absent from the young adults scored to construct Figure 1). Low down in the ovarioles

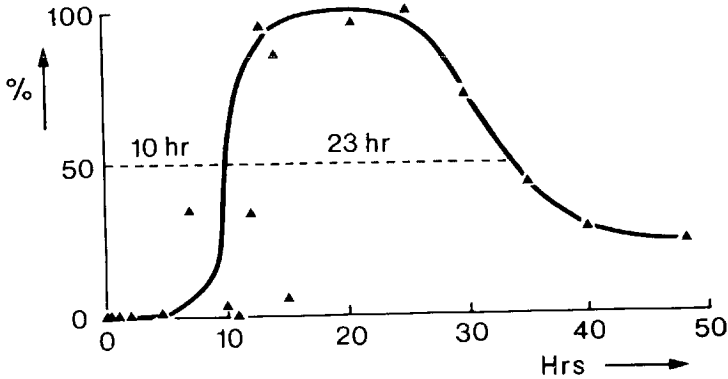


Figure 1. Percentage labelled metaphases at 32°C in somatic cells of the ovarian follicles of young, adult *M. virgo* versus time after injection with ^3H -thymidine. Colcemid was not used. Percentages include 30-200 metaphases.

the follicle cells stopped dividing but continued to replicate, increasing the nuclear size dramatically. At 32°C the average G2 in spermatogonial cells was about 12 hours and about 4 hours in gastric caeca, but both were extremely variable. In insects used primarily to study ovarian follicle and spermatogonial divisions, metaphases in the gastric caeca were usually mid-S or early-S labelled, but the G2 was sufficiently variable to allow a few to be late-labelled. A high variation in the cycle times of individual cells has been found in simple tissues in culture (Comings, 1970) and these times must be expected to be at least as variable in complex tissues in our *in vivo* experiments on morabines. In experiments primarily designed to show late-labelling in the gastric caeca over 90% of the ovarian follicle metaphases were unlabelled and the rest were late-labelled. Only a small proportion of cells in the gastric caeca were diploid and undergoing division; the majority of cells were large (Figure 2) and able to replicate but not to divide. It has been shown that different somatic tissues have similar labelling patterns in some animals (e.g. Martin, 1966). The patterns of labelling in morabine gastric caeca and ovarian follicle cells are very similar and although there is a great difference in length of the average G2 in both cell classes the follicle cells behave as if they were somatically derived (Chapman, 1969). The labelling patterns of spermatogonial cells are varied and remain somewhat obscure but certain features can be related to the patterns in somatic cells.

It has been generally found that heterochromatin replicates

faster and later than euchromatin so that the time lag in initiation of replication of the heterochromatin after euchromatin is longer than the lag in termination of replication. In a study involving flash labelling of certain members of the *cultrata* group of morabine grasshoppers (injection of tritiated thymidine followed by sacrifice 30 min later) the metaphases were unlabelled and the relative number of diploid interphase nuclei labelled at different stages of S-phase indicated faster replication of heterochromatin than euchromatin (Table 1; any bias in the classification of the cells would have increased the score of mid-S labelled cells and decreased the early S score).

TABLE 1

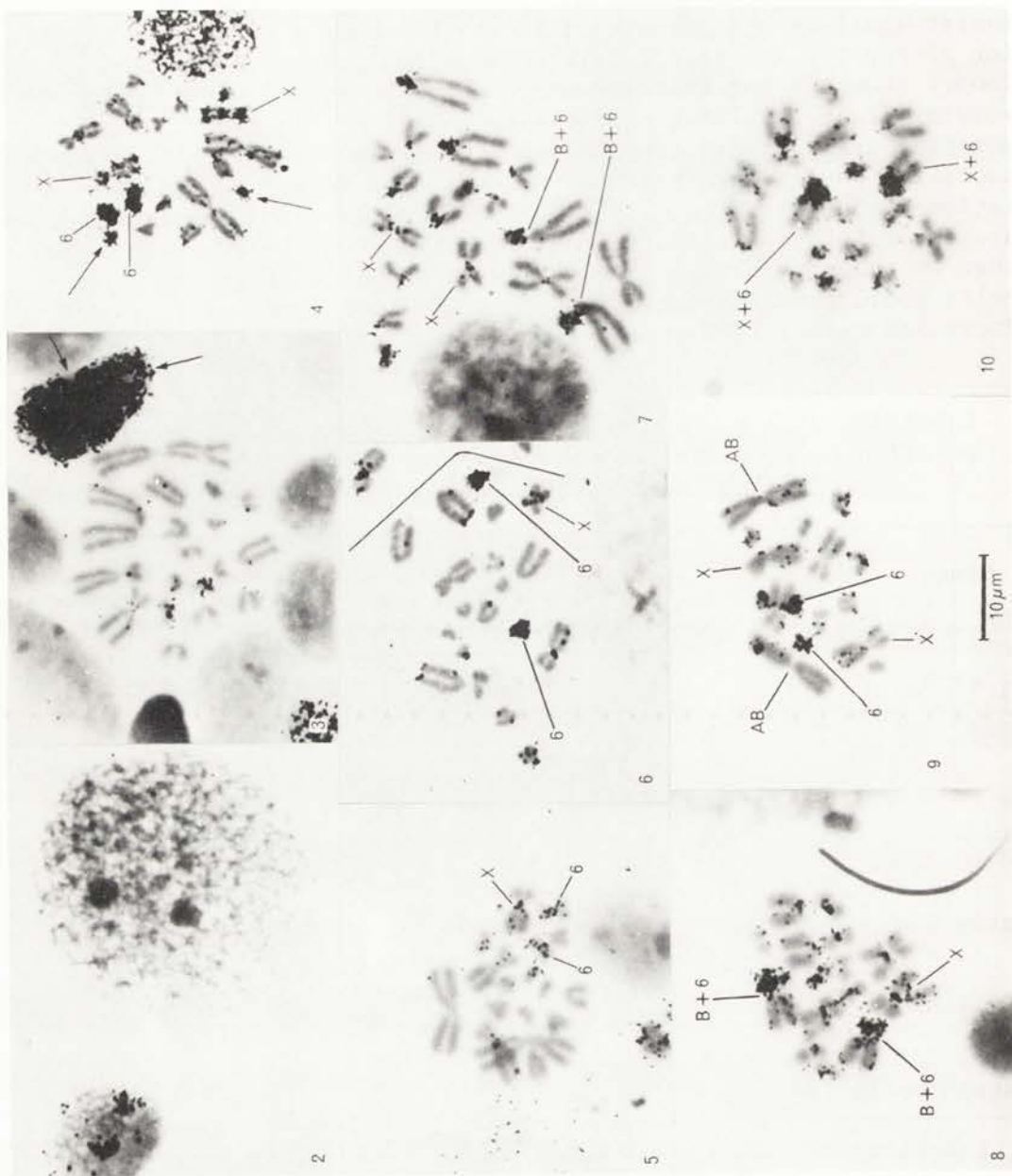
Labelling of diploid interphase nuclei after $\frac{1}{2}$ hour incorporation of tritiated thymidine: scores (and percentages) of cells in early, mid and late S-phase

Stage	Labelling patterns		♂ Gastric caeca	♀ Gastric caeca	Ovarian follicles
	Het.	Euch.			
G ₁ or G ₂	-	-	1,369	1,831	1,711
Late	+	-	17 (19.5)	27 (20.8)	22 (19.0)
	++	+			
Mid S	+	+	39 (42.5)	50 (38.5)	37 (31.9)
	++	++			
Early S	-	+	33 (37.9)	53 (40.8)	57 (49.1)
	+	++			
Total			87 (100.0)	130 (100.0)	116 (100.0)
Early S/ late S cells			1.94	1.96	2.59

Labelling density: - no label, + light label, ++ heavy label

LABELLING PATTERNS IN AUTOSOMES

The most prominent heterochromatic structures in somatic interphase nuclei are a pair of small autosomes which may be called the megameric pair (Figure 2). Megameric chromosomes appear to be fundamental components of the karyotypes of the Orthoptera. The replication of the heterochromatin of the megameric pair begins and



Figures 2-10. All chromosomes are late-labelled. Figures 2 and 4-5 are from species 'P44a'. Figure 3 is from species 'P151'. Figures 6-9 are from members of the *viatica* group of species and races. Figures 3, 4, 7, 9, 10 are ovarian follicle metaphases. Figures 5, 6, 8 are spermatogonial metaphases. Figure 2. A large nucleus from the gastric caeca of a female showing two heterochromatic bodies and a smaller diploid nucleus late-labelled mainly over the heterochromatin. Figure 3. Labelled very late in S-phase with most grains over the megameric pair of chromosomes. A heavily labelled interphase is early-labelled with heterochromatin unlabelled (arrows). Figure 4. Late-labelled but earlier than in Figure 3. The megameric no. 6 pair and the X chromosomes are labelled and a small pair of late-labelling autosomes are arrowed. Figure 5. The X, the no. 6 pair and a small pair labelled. Figure 6. From the race of the species '*viatica*' with 19 chromosomes in the male. The heavily labelled, megameric no. 6 pair are independent. Figure 7. From the 17 chromosome race of '*viatica*' in which the megameric no. 6 is fused to the large autosomal pair (B). The short arms of acrocentric chromosomes are relatively long in this species and are late-labelling. Figure 8. From species 'P24XY' which has the B+6 fusion as in Figure 7. The autosomally-derived region of the X is not so late-labelling as the region derived from the X in species with XO males. Figure 9. From the translocation race of 'P24XY'. The translocation which has formed a large AB pair of autosomes has released the megameric no.6 pair. Figure 10. From the species 'P45bXY' in which the X has become fused in a tandem fashion to the distal end of the megameric no. 6 pair.

terminates later than the euchromatin of the other chromosomes although at the end of S-phase the observed differences in labelling are probably due more to differences in the rate of replication than in the time of termination. However, it is usually possible to find a few interphases and metaphases which demonstrate completely labelled heterochromatin or megameric chromosomes against a background of unlabelled euchromatin or non-megameric chromosomes (Figure 3).

In most species the megameric pair are late to begin and end replication in spermatogonial divisions but the degree of late replication is reduced by comparison with somatic cells (Figures 4 and 5). The megameric chromosomes are usually independent but in some species and races of the *viatica* group of morabine grasshoppers (White et al., 1969) they have become fused, either to other autosomes or to the X chromosomes and in one case they have been freed from fusion by a subsequent translocation (Figures 6-10).

In many species a second pair of small autosomes are also relatively late-labelling but less so than the megameric pair (Figure 4). This other pair do not form visible heterochromatin at interphase. In cells labelled late but not at the very end of S-phase the centromeric regions of most chromosomes tend to be

labelled and there is a marked increase of this labelling in forms which have measurable short arms on the acrocentric chromosomes (Figure 7). Centric heterochromatin is probably present and the increase in the length of the short arms appears to be due to the addition of heterochromatic segments which appear condensed at leptotene but not in somatic nuclei.

LABELLING AND HETEROPYCNOSIS IN THE SEX CHROMOSOMES

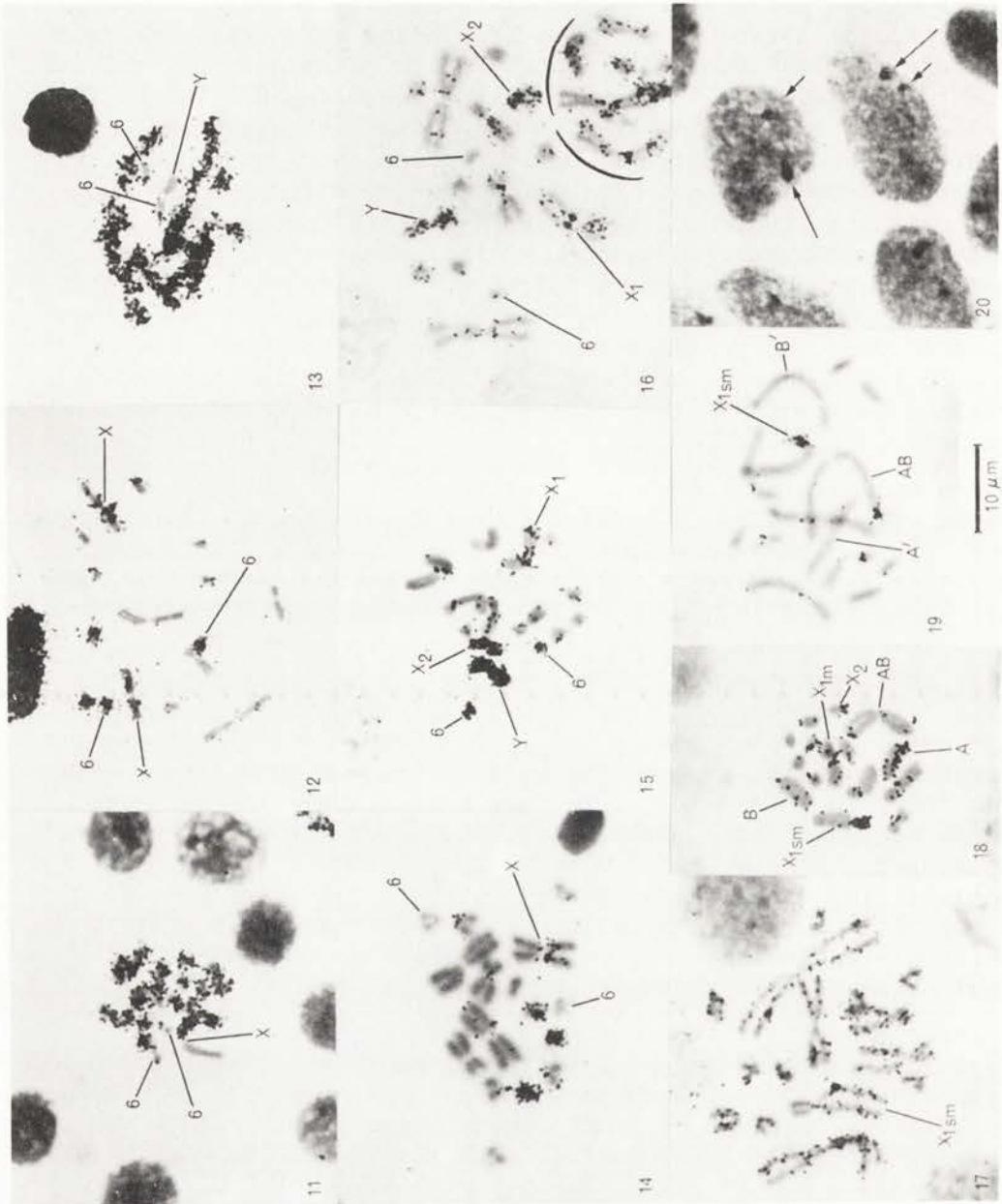
Negative heteropycnosis of the X chromosome during early spermatogonial divisions which morabines share with other short-horned grasshoppers is diminished with the use of colcemid but is still visible. At interphase the X chromosome is in a separate region of the nucleus and it remains on the periphery of the spread chromosomes at metaphase, being easily lost on squashing. The X chromosome is usually only moderately late-labelling in spermatogonial divisions (Figure 6). However, the whole of the X in species with XO males, or part of the X in species with XY males, may be spectacularly underlabelled in spermatogonial divisions labelled in early S-phase (Figure 11). Results are very variable in cells progressing through the spermatogonial series. In somatic cells of both sexes the X chromosomes are moderately late replicating to varying degrees in different species (Figures 4-10,12,15,16) but they are not notable laggards in initiating replication. The X chromosomes do not form condensed somatic interphase heterochromatin. They are consequently in a separate class of late-labelling chromosomal regions which also includes some autosomal segments that show moderately late termination of replication, but not the late initiation and interphase heterochromatinization shown by the megameric pair in somatic cells.

In the *cultrata* group of species, forms with XO, XY and X_1X_2Y males are found. Three of the species with XO males investigated by autoradiography had a basic morabine karyotype exhibiting all of the labelling properties discussed previously. The megameric chromosomes are metacentric and independent in all members of the group investigated but they vary considerably in size between species. In *cultrata*, one of the XO species, the megamerics are large and very heavily labelled late in S-phase in somatic cells (Figure 4) and as usual somewhat less so in spermatogonial cells (Figure 5). On the other hand the two species with XY and one species with X_1X_2Y males (White and Cheney, 1966) have very small megameric chromosomes which are late-labelling in somatic cells but with fewer grains than the most heavily late-labelled small autosomal pair (Figure 12). In the two species with XY males the Y chromosome is late-labelling in somatic cells as well as being almost as late in initiating replication as the megameric pair (Figure 13); it forms a somewhat less

condensed interphase heterochromatin body than the latter. In spermatogonial divisions in these XY species the megameric chromosomes are not late replicating and in fact are among the earliest pairs to complete replication. Thus in spermatogonial metaphases labelled very late in S-phase the Y and one small pair are labelled (Figure 14) and at a slightly earlier stage the X chromosome is moderately late-labelled. In spermatogonial divisions the megameric pair also appear to begin replication later than the other autosomes but metaphases uniformly labelled at mid S-phase can be found and this must be the period when the megamers are replicated. This complete reduction in late-labelling ability of the megameric pair during spermatogonial division appears to be an advance on the reduction of the degree of late-labelling in the same divisions in species with larger megameric chromosomes with more heterochromatic material.

The moderate somatic late-labelling of the X chromosome in species with XO males is reduced in the XY species of the *cultrata* group. In the two races of an X_1X_2Y species the X chromosome, equivalent to the X in the XY species, is most late-labelled over short segments, distal on one arm and proximal on the other. The megameric chromosomes are small and behave as in the XY species. The X_2 and Y chromosomes, which were autosomal prior to the origin of the X_1X_2Y system, are very late-labelling in all types of cells (Figures 15 and 16). The X_2 and Y also form clear interphase heterochromatin and with the two very small bodies derived from the megameric pair, there is a total of four heterochromatic bodies in somatic nuclei in both sexes. In the *cultrata* group of morabine grasshoppers there appears to be an upper limit to the amount of late-labelling material in the nuclei. The labelling behaviour of the megameric chromosomes indicates that they may have a critical function in spermatogenesis which probably accounts for their universal occurrence in grasshoppers.

Since the majority of cells in insect somatic tissues are endopolyploid, regulatory mechanisms in polyploid nuclei are particularly important in insects. We have made some observations on heteropycnosis in tetraploid cells of one morabine species which may have a bearing on this general question (White, 1970). The species in question has a neo-XY sex chromosome mechanism, the longer limb of the X (X_q) being the original X chromosome, the shorter one (X_p) being of autosomal origin. In the normal diploid spermatogonia X_q shows the usual negative heteropycnosis in the early cell generations. In one individual, however, two cysts of tetraploid spermatogonia had arisen spontaneously (256 cells altogether). Although the two X_s in every nucleus must have been molecular copies of one another,



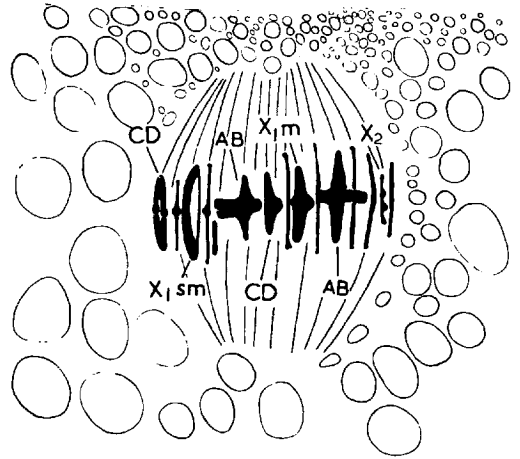
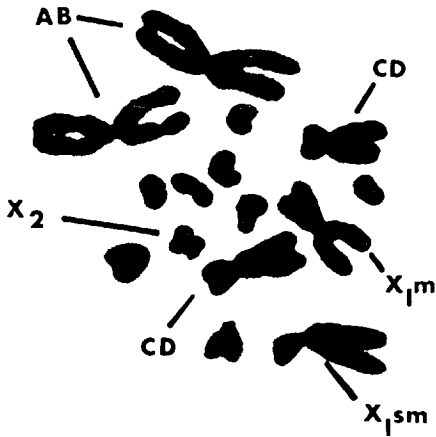
Figures 11-20. Figures 11-16 are from members of the *cultrata* group of morabine grasshoppers. Figures 11-14 are from a species with XY males, Figures 15 and 16 from a species with X_1X_2Y males. Figures 17-20 are from *M. virgo*. Figures 11, 14 and 16 are spermatogonial metaphases. Figures 12, 17, 19 and 20 are ovarian follicle metaphases or nuclei. Figures 13, 15 and 18 are metaphases from gastric caeca. Figure 11. Early-labelled with one arm of the X and the megameric no. 6 pairs unlabelled. Figure 12. Late-labelled with grains over the X, a pair of small autosomes and the megameric pair (identified by removal of the grains). Figure 13. Early-labelled with the Y and the no. 6 pair underlabelled. Figure 14. Late-labelled with three small chromosomes, which must be the Y and two small acrocentric chromosomes, well labelled. The no. 6 pair are unlabelled and negatively heteropycnotic. Figure 15. Late-labelled with many grains over the X_2 , Y and no. 6 pair. Figure 16. Late-labelled with most grains over the X_2 and Y. Figure 17. Early-labelled with the limb of the X sm underlabelled and showing close chromatid apposition. Standard karyotype. Figure 18. Late-labelled Monia Gap karyotype. Figure 19. Labelled very late in S-phase, Yatpool karyotype. Figure 20. Interphase nuclei in ovarian follicle cells of *M. virgo* showing unequal heterochromatin (long arrow, larger body, short arrow, smaller body).

only *one* of the two Xq limbs manifested negative heteropycnosis, the other being condensed to the same extent as the autosomal euchromatin. The asymmetry of heteropycnosis of the two Xq limbs could be seen in prophase, metaphase and anaphase-telophase. Since this case was discovered accidentally, no autoradiographic studies on the DNA replication pattern were carried out, but it seems clear that the two Xq limbs must have been asynchronous for the timing of replication as well as for the cycle of condensation.

Clearly this asynchrony of chromosomal behavior, whatever its consequences for gene action in the two Xq limbs, is not itself an adaptive genetic mechanism (such as the inactivation of one X in the soma of female mammals undoubtedly is), since the tetraploid spermatogonia are an abnormal phenomenon and cannot give rise to normal sperms. It may, however, be a reflection of some regulatory process which occurs normally in the endotetraploid, endooctoploid, etc. nuclei of the soma.

LABELLING PATTERNS IN *MORABA VIRGO*

The parthenogenetic all-female species *Moraba virgo* is an unique case of a diploid organism which is extensively heterozygous for late-replicating segments in several chromosome pairs. *M. virgo* usually has $2n = 15$ (Figure 21). However, synapsis produces 15 bivalents so the chromosome number must have been doubled at a premeiotic stage (Figure 22). But synapsis is strictly confined to sister chromosomes which are molecular copies of one another, having been produced by an extra premeiotic replication.



Figures 21- 22. The Standard karyotype of *M. virgo*, $2n = 15$, found at most of the known localities. Figure 21. Mitotic metaphase from an ovarian follicle cell. Figure 22. Camera lucida drawing of first metaphase in the egg showing 15 bivalents.

We know that this is so, because the standard karyotype of *virgo* is heterozygous for various structural changes and at a number of localities the individuals show karyotypes which are even more highly heterozygous, since they have various fusions or translocations superimposed on the standard karyotype. In spite of this, no structurally heterozygous bivalents are ever observed at meiosis, neither do multivalents occur.

The fifteen bivalents of the first meiotic division show perfectly typical chiasmata, but these clearly do not lead to genetic recombination, since they are between chromatids which are strictly identical. There are two meiotic divisions which are quite orthodox, except that there are twice as many bivalents as one would have expected in an organism with 15 chromosomes in the somatic karyotype. It will be evident that with this genetic system all the offspring of a given female should be genetically identical to one another and to their mother. Moreover this is a system in which any genetic change that establishes itself in evolution will be incorporated in the genotype in the heterozygous state and can never become homozygous (except in the very unlikely event of an identical change occurring by spontaneous mutation in the homologous chromosome).

The 'standard' karyotype of *M. virgo* has been found at approximately 30 localities in New South Wales, Victoria and Western Australia. Three other karyotypes will be referred to as the Monia Gap, Yatpool and Hayes Hill karyotypes. The standard karyotype includes a pair of large equal-armed metacentric autosomes (the AB pair), a pair of X_1 chromosomes, a pair of unequal-armed autosomes

(the CD pair), four pairs of small acrocentric autosomes and a small unpaired (monosomic) metacentric X_2 chromosome. The designation of the sex chromosomes as X_1 and X_2 is due to the fact that the bisexual ancestor of *M. virgo* was probably a form with an X_1X_2Y (δ) : $X_1X_1X_2X_2$ (♀) sex chromosome constitution. That ancestral species or population has not been found, and is presumably extinct, but a closely-related species ('P196') with X_1X_2Y males and $X_1X_1X_2X_2$ females exists (White et al., 1973) and has even been successfully hybridized with *virgo* in the laboratory. The evolutionary history of the sex chromosome mechanism in the lineages leading to *virgo* and 'P196' has been complex, but is reasonably well understood. The X_1 chromosomes of *virgo* include the original X of a remote ancestor which was XO (δ) : XX (♀) and a small autosome which fused with it at an early stage in the evolution of *virgo* (i.e. before parthenogenesis had been established as the mechanism of reproduction. X_2 , which is hemizygous in *virgo* and in the male of 'P196' (but diploid in the female of the latter species) was likewise an autosome in the past.

The two X_1 chromosomes of *virgo* are structurally different. One of them is an equal-armed metacentric (X_{1m}) while the other, which may have a slightly greater overall length, is a submetacentric (X_{1sm}), with one limb about three times the length of the other. This difference between the two X_1 chromosomes, which exists in every individual of *virgo*, probably arose by a pericentric inversion, but may well have involved other structural changes as well, it probably arose in 'bisexual times', before the switch to parthenogenetic reproduction. It is necessary to point out that in the early period of our work on *M. virgo*, before the related bisexual species had been examined, we wrongly identified the CD pair as X chromosomes, the X_1 chromosomes being regarded as a pair of CD autosomes (White and Webb, 1968). This error of interpretation was only corrected when the discovery of 'P196' proved that *virgo* had arisen from a bisexual species with X_1X_2Y males and enabled us to understand the nature of the little X_2 chromosome, previously called m_2 . There is, however, still no explanation for the fact that X_2 is hemizygous in *virgo*.

We have studied the DNA replication pattern in somatic cells of *M. virgo*, using the standard techniques of tritiated thymidine autoradiography on ovarian follicle cells and the cells of the gastric caeca. These gave substantially the same result, provided one takes account of the fact that the G2 period is much shorter in the caecal cells. The general pattern of DNA labelling seems to be the same in material from all localities; in particular it appears to be substantially the same in material from New South Wales and Western

Australia, although the possibility of some minor differences has not been definitely excluded.

In metaphases which we interpret as having been labelled very early in the S-phase the chromosomes appear to be fairly uniformly labelled, except for the fact that the X_{1sm} p limb has a very low grain count and may be totally unlabelled. Usually the two chromatids of this limb show the phenomenon of close apposition (well known in the case of the human Y chromosome) at this stage, that is, they do not diverge from one another at an angle under the influence of the colcemid, as do the chromatids of the other chromosome limbs (Figure 17).

In metaphases labelled late in the S-phase the most noteworthy feature is the very large number of grains over the X_{1sm} p limb, which is now seen to be strongly late-labelling (Figure 18). The two AB chromosomes consist of one homolog (the 'high label AB'), which shows a higher grain count over one arm (which we call the A limb) than over the other (the B limb), and a 'low label AB' both of whose limbs show a lower grain count than the B limb of the other homolog. The low label AB shows a fairly uniform low grain count over both limbs except for a short distal segment on one limb which shows quite a high count. The grain counts over the two CD chromosomes do not seem to be significantly different and the grain counts of the eight short autosomes and the X_2 do not call for special comment.

In metaphases labelled very late in the S-phase most of the chromosomes are unlabelled. Three chromosomal structures still show considerable grain counts at this stage: the X_{1sm} p limb, the extreme tip of one of the limbs of the 'low label AB' (Figure 19), and one of the eight small acrocentric autosomes. Probably the second of these is normally the last region of the karyotype to complete replication.

In material with the Monia Gap and Yatpool karyotypes, in both of which one of the AB chromosomes has been involved in a translocation (a different translocation in each case), it can be seen that the translocated AB is in both instances the high label one, which has been broken in the vicinity of the centromere. Following the translocation the A and B limbs retain their characteristic labelling patterns. These observations on karyotypes with translocations involving the AB prove that the asymmetry in replication between the two homologs is an indication of a basic and permanent difference between them. It is hence not analogous to the different behaviour of the two Xs in the somatic cells of female eutherian mammals, where the paternal X is late-labelling in some cells and the maternal one in others.

It is evident that *M. virgo* is a permanent heterozygote for labelling pattern in at least three of its seven pairs of chromosomes. Minor differences in the labelling pattern of the other four pairs (the CD pair and three pairs of small acrocentrics) may exist but only relatively gross differences are detectable by the technique of counting grains. It is necessary to point out that we have carried out extensive statistical tests on our grain count data which demonstrate that the differences between the AB homologs (and between the two limbs of the high label AB) are highly significant (White and Webb, 1968). *Moraba virgo* does not show a pair of late-labelling 'megameric' small autosomes in its karyotype. Since such a pair is present in a related species 'P151', with XO males (Figure 3), it is fairly clear that the megameric pair have disappeared in *virgo*, perhaps being incorporated in the X_{1sm} p limb.

Various types of tissues in *M. virgo* show interphase nuclei in which two unequal heterochromatic masses can usually be seen (Figure 20). The larger of these is almost certainly the X_{1sm} p limb, while the smaller may be the tip of one limb of the 'low label' AB. The condensation of these regions in the somatic nuclei must be accepted as a strong indication that they are genetically inactivated or 'switched off'. Whether the less strongly marked differences in labelling pattern between the remaining segments of the X_1 and AB pairs indicate inactivation of some loci or blocks of loci, in one member of each pair, is less certain but appears probable. If this interpretation is correct, *M. virgo*, although cytologically a diploid, would have evolved some way along the path leading to 'functional haploidy'. On the other hand, there are two lines of evidence showing that loss of unneeded functions has not occurred to any great extent since the loss of sexual reproduction. One is that all the structures required for copulation, including the spermatheca, are fully developed in *virgo*. More remarkable is the finding that *virgo* still carries genes capable of affecting the shape of the male subgenital plate in triploid hybrids between it and related bisexual species which have been obtained in laboratory crosses. Thus, whatever the correct interpretation of the extensive development of heterozygosity for late-replicating DNA segments may be, it probably has nothing to do with loss of functions which would no longer be required in an all-female species.

Clearly, fixed heterozygosity for late-replicating segments, of the type found in *M. virgo*, could not exist in a species with effective crossing over at meiosis. It is a genetic system which is likely to be found in other parthenogenetic species which lack genetic recombination.

ACKNOWLEDGMENT

Supported by Public Health Service Grant GM-07212 from the Division of General Medical Sciences, U.S. National Institutes of Health and by a grant from the Australian Research Grants Committee.

REFERENCES

- Chapman, R.F. 1969. The Insects, Structure and Function. English University Press, London.
- Comings, D.E. 1970. Quantitative autoradiography of heterochromatin replication in *Microtus agrestis*. *Chromosoma* 29: 434.
- Martin, P.G. 1966. The pattern of autosomal DNA replication in four tissues of the Chinese hamster. *Exp. Cell Res.* 45: 85.
- White, M.J.D. 1970. Asymmetry of heteropycnosis in tetraploid cells of a grasshopper. *Chromosoma* 30: 51.
- White, M.J.D. and J. Cheney. 1966. Cytogenetics of the *cultrata* group of morabine grasshoppers. I. A group of species with XY and X_1X_2Y sex chromosome mechanisms. *Aust. J. Zool.* 14: 821.
- White, M.J.D. and G.C. Webb. 1968. Origin and evolution of parthenogenetic reproduction in the grasshopper *Moraba virgo* (Eumastacidae: Morabinae). *Aust. J. Zool.* 16: 647.
- White, M.J.D., K.H.L. Key, M. André and J. Cheney. 1969. Cytogenetics of the *viatica* group of morabine grasshoppers. II. Kangaroo Island populations. *Aust. J. Zool.* 17: 313.
- White, M.J.D., G.C. Webb and J. Cheney. 1973. Cytogenetics of the parthenogenetic grasshopper *Moraba virgo* and its bisexual relatives. I. A new species of the *virgo* group with a unique sex chromosome mechanism. *Chromosoma* 40: 199.

Studies of the molecular anatomy of the SV40 viral genome

Paul Berg, Janet E. Mertz and John Carbon

The *EcoRI* and *HpaII* restriction endonucleases each cleave Simian Virus 40 (SV40) DNA once at map positions 0 and 0.735 SV40 fractional length, respectively. Specific mutants whose DNAs are resistant to cleavage by these two enzymes have been isolated. One class (10 mutants), containing deletions varying from 80 to 190 base pairs and including the *HpaII* restriction sequence, grows without a helper virus although more slowly than wild-type. Viral mutants with a similar phenotype have been synthesized *in vitro*. In the procedure, about 50 (dA:dT) base pairs are inserted at the *HpaII* restriction site by cyclizing and covalently sealing *HpaII*-cut SV40 linear DNA molecules containing a short 'tail' of poly(dA) residues at one 3' end and of poly(dT) residues at the other 3' end. Since insertion of a new sequence or deletion of an existing sequence at map position 0.735 alters, but does not prevent viral multiplication, this region either has a regulatory function or it codes for a dispensible portion of one of the late proteins. Although most (>99%) of the *HpaII*- or *EcoRI*-resistant defective SV40 DNA obtained by high multiplicity infections fails to produce plaques, 'early'-antigen synthesis and viral DNA replication occur nearly normally. Individual clones of defective SV40, lacking the *EcoRI*- and/or *HpaII*-restriction sequences, have been isolated and propagated using SV40 temperature-sensitive mutants as helpers. By analyzing heteroduplexes formed between mutant and parental DNAs and restriction endonuclease-generated fragments, the nature and locations of the modifications in the DNA structures can be deduced. These mutants contain extensive deletions within the region of 0.72 to 0.15 SV40 map unit and, generally, sizable duplications of the region containing the origin of DNA replication.

Viruses bring new genes into the cells they infect. Generally the new genetic information serves to establish the machinery for multiplying the virus, specifically, the precursors and proteins needed to replicate and encapsidate the viral nucleic acids. With certain viruses and their appropriate hosts the infection has an alternative outcome. The infected cell survives and occasionally undergoes a heritable alteration of its morphology and growth characteristics, particularly the ability to grow under conditions where normal cells are arrested and to initiate tumors in appropriate animals. Such transformed cells invariably contain and transmit at each cellular division all or part of the viral genome as an integral part of the cell's chromosomal DNA; moreover,

expression of one or more of the integrated viral genes is probably required to maintain the transformed phenotype.

Each of these responses, viral multiplication and cellular transformation, can be elicited with the small DNA virus, Simian Virus 40 (SV40). Because SV40 contains only a very limited amount of genetic information - 5,100 base pairs or enough coding capacity for about five proteins of 40,000 daltons molecular weight - one can be optimistic at the prospects of understanding the genetic control and molecular events of both the multiplication cycle and the transformation phenomenon.

What are these SV40 viral genetic elements and how are they organized physically on the viral DNA molecule? Furthermore, what are the viral gene products, how do they function and what is the arrangement of viral genes after integration into the cellular DNA? These are formidable questions and undoubtedly their answers will be some time in coming. It seems logical, however, to begin with the first question.

Several laboratories (Robb and Martin, 1970; Tegtmeier and Ozer, 1971; Kimura and Dulbecco, 1972; Kit et al., 1970; Chou and Martin, 1974) have isolated and characterized conditional-lethal, temperature-sensitive (*ts*) mutants of SV40. Although such mutants can provide valuable information about the physiology and molecular biology of the viral life cycle (Robb and Martin, 1972; Robb et al., 1972; Tegtmeier, 1972), their utility is limited to genes that code for proteins and by the relative difficulty in accurately mapping the mutant loci on the viral DNA molecule. On the other hand, substantial alterations in the viral DNA structure (e.g. deletions, duplications, insertions and substitutions) would very likely alter the normal viral phenotype and could be mapped by heteroduplex analysis (Morrow and Berg, 1972; Davis et al., 1971). This paper summarizes our general approach for the isolation, construction and propagation of such grossly defective SV40 mutants and presents a preliminary characterization of several representative defective mutants.

Coordinates for the SV40 DNA map

Since the SV40 genome is contained in a continuously circular, that is, covalently closed double-stranded DNA molecule, one or more specifically located reference points are needed to fix the position of a particular genetic locus or region. Bacterial restriction endonucleases, enzymes which make double-strand cleavages in DNA molecules at specific nucleotide sequences, serve admirably to provide such coordinates (Figure 1). A number of enzymes have proven to be quite useful for this purpose.

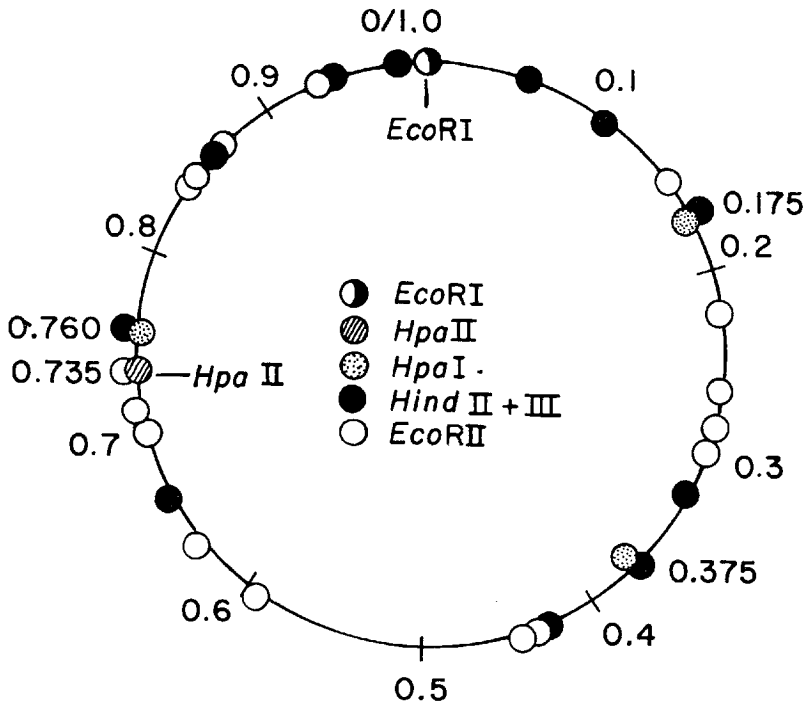


Figure 1. Sites of cleavage of SV40 DNA by several restriction endonucleases.

a) *Eco*RI (Smith and Nathans, 1973) restriction endonuclease (Yoshimori, 1971) cleaves SV40 DNA once to produce a non-permuted population of unit length linear molecules (Morrow and Berg, 1972; Mulder and Delius, 1972; Hedgpeth et al., 1972). We have defined the *Eco*RI restriction site as position 0 of the SV40 DNA map (Morrow and Berg, 1972) and the coordinates are expressed as SV40 DNA fractional lengths.

b) Another useful enzyme, from *Hemophilus parainfluenzae*, is *Hpa*II restriction endonuclease (Sack and Nathans, 1973; Sharp et al., 1973). This enzyme also produces only one double-strand scission per SV40 DNA molecule at a position 0.735 SV40 fractional length, clockwise, from the *Eco*RI cleavage site (0.735 SV40 map unit) (Sharp et al., 1973; Morrow and Berg, 1973).

c) *Hpa*I, a second enzyme from *H. parainfluenzae*, cleaves SV40 DNA at three places and these have been mapped at 0.175, 0.375 and 0.760 SV40 map unit (Sack and Nathans, 1973; Sharp et al., 1973).

d) Several other restriction endonucleases cut SV40 DNA more extensively. A mixture of at least two enzymes from *Hemophilus influenzae* (*Hind* II + III) cleaves SV40 DNA at eleven locations (shown on the map by black dots) to yield eleven distinguishable

fragments (Danna and Nathans, 1971). Another enzyme, *EcoRII* coded for by the RII plasmid in *Escherichia coli* (Yoshimori, 1971), makes sixteen cleavages in SV40 DNA, one at each of the open circles on the map (Subramanian et al., 1974).

The availability of these enzymes and knowledge of their cleavage sites has permitted the selection, construction and characterization of SV40 mutants with alterations in the wild-type DNA structure.

Individual clones of defective, helper-dependent SV40 virus

SV40 virus particles with defective genomes accumulate when plaque purified virus is serially passaged in susceptible cells at high multiplicities of infection (moi) (Yoshiike, 1968; Yoshiike and Furuno, 1969; Tai et al., 1972). These defective viral DNAs are generally smaller than the DNA obtained from plaque purified virus and have gross alterations (deletions, substitutions, additions) in their molecular structure (Yoshiike and Furuno, 1969; Tai et al., 1972). Since the molecular defects contained in these DNAs are distributed somewhat randomly throughout the molecule (Mertz and Berg, unpublished), these preparations provide a rich source of mutants with substantial modifications in various portions of their genomes.

Our starting material for the isolation of such mutants was obtained by four serial passages of wild-type SV40 strain Rh9.11 (Girardi, 1965) (*WT800*) on primary African Green Monkey kidney cells using undiluted lysate obtained from each passage to initiate the subsequent infection. The SV40 DNA obtained from the fourth cycle of infection was judged to be grossly defective by the following criteria: the titer (plaque forming units (pfu)/ml) was more than two hundred-fold lower than that of an analogous lysate produced by low moi infection; the supercoiled DNA from the infected cells was very heterogeneous with molecules ranging in size from about 0.6 to somewhat greater than 1.0 SV40 fractional length; the defective DNA yielded deletion, substitution and addition loops when heteroduplexes with wild-type (*WT*) SV40 DNA were examined by electron microscopy; and, whereas about 99% of *WT* DNA preparations are cleaved once at map positions 0 and 0.735 by the *EcoRI* and *HpaII* restriction endonucleases respectively, 55% of the DNA was resistant to cleavage by *EcoRI* endonuclease, 47% was resistant to *HpaII* endonuclease and 23% was resistant to both enzymes (Mertz and Berg, unpublished).

The *EcoRI*-resistant and *HpaII*-resistant DNAs were separated from the molecules which were cut by the enzymes by equilibrium centrifugation in cesium chloride-ethidium bromide gradients (Radloff

et al., 1967) and then used as the sources from which cloned isolates of defective SV40 were recovered.

Preparations of *EcoRI*- and *HpaII*-resistant DNAs were four orders of magnitude less infectious than *WT* DNA in conventional plaque assays. Consequently, to isolate and propagate such defective genomes, a procedure which permits complementation of their defect was required. Temperature-sensitive (*ts*) mutants of SV40 can be used for this purpose. To decide which of the known *ts* mutants would complement the growth of these defective DNAs, it was necessary to identify the defective function(s) in the *EcoRI*- and *HpaII*-resistant DNA populations. Following infection, the *EcoRI*- and *HpaII*-resistant DNAs induce the 'early' SV40 specific T antigen and replicate their DNA nearly normally, but they fail to produce the 'late' virion capsid antigen. Consequently, molecules lacking either the *EcoRI*- or *HpaII*-restriction site (map position 0 and 0.735, respectively) can express those functions preceding and including DNA replication ('early' functions) but not the 'late' functions; this result agrees with previous reports (Morrow et al., 1973; Sambrook et al., 1973; Khoury et al., 1973) that these two restriction sites are contained within the 'late' region of the SV40 genome. On this basis *tsA30*, a non-leaky 'early' mutant which can complement the 'late' mutants *tsB4* and *tsB11* for growth (Tegtmeyer and Ozer, 1971) was selected as the 'helper' virus to grow the restriction enzyme-resistant mutants. (These SV40 *ts* mutants were isolated and very generously sent to us by P. Tegtmeyer.)

To isolate individual clones of defective SV40 lacking the *EcoRI*- and *HpaII*-restriction sites, we coinfecting monkey cells with the restriction endonuclease-resistant DNA preparations and *tsA30* DNA, and then incubated the cells at 41°, the restrictive temperature for *tsA30* growth. Although some of the plaques that appear were produced by a trace of *WT* molecules contaminating the defective DNA population or by viable defectives (see below), the majority of the plaques resulted from cells doubly infected with defective and complementing helper genomes. These putative mixed plaques were picked and plaque-purified by successive coinfections at 41° with added *tsA30* helper virus. Generally, after one or two plaque-purifications, the virus suspension obtained from a mixed plaque produced 10 to 100-fold more plaques at 41° in the presence of added *tsA30* virus than in the absence of the helper; and, whereas the former occurs with single-hit kinetics, the latter more nearly approaches a two-hit process.

Eleven independent plaque isolates, five from the *EcoRI*-resistant DNA pool and six from the *HpaII*-resistant DNA pool, were selected for further study. DNA preparations and high titer virus stocks of these putative defective mutants were obtained by infection

of monolayers of CV-1P cells with a mixture of *tsA30* virus and virus from the last mixed plaque and incubation at 41°. At the appropriate time viral DNA or virus was isolated from the cultures by standard methods.

To obtain pure defective DNA from each of the clones, the mixed pool of closed circular DNA recovered from the mixed infection was incubated with *EcoRI* restriction endonuclease (if it was derived from the *EcoRI*-resistant DNA pool) or with *HpaII* restriction endonuclease (if it came from the pool of *HpaII*-resistant DNA) and the restriction endonuclease-resistant material from each isolate was then separated from the linear helper DNA by velocity sedimentation. For the eleven clones we have isolated the endonuclease-resistant or defective DNA represented between 50 to 96% of the total DNA recovered; with ten however, more than 75% of the DNA was of the defective mutant.

Electron microscopic examination of the restriction endonuclease-resistant DNA fraction from each plaque isolate revealed essentially homogeneous populations of circular molecules (Table 1); the mean sizes of these mutant DNAs ranged from 0.71 to 0.96 SV40 fractional length. In a subsequent publication (Mertz and Berg, in preparation) we shall report a more detailed analysis of these defective DNAs; these experiments show that the defective DNAs lack either the *EcoRI*- and/or *HpaII*-restriction endonuclease cleavage sites because they have extensive deletions of DNA encompassing these regions of the genome. Moreover, most of the mutants also have sizable duplications of the region of the SV40 genome that includes the origin of DNA replication (Nathans and Danna, 1972; Fareed et al., 1972) which probably accounts for their preferential growth in mixed infection with the helper. In this paper we describe the characterization of one of these mutants, *dl-812* (Table 1), to illustrate how this analysis was done and what deductions concerning the structure and origin of the defective DNA can be drawn.

The DNA of mutant *dl-812* is: a) 0.11 SV40 fractional length shorter than wild-type DNA; b) resistant to cleavage by *EcoRI* endonuclease; and c) cleaved once by *HpaII* endonuclease. Heteroduplexes formed after denaturation and renaturation of a mixture of *HpaII* endonuclease-cleaved mutant and wild-type linear DNA (suitable single- and double-stranded marker DNAs were also added to serve as internal length standards) were mounted and examined by electron microscopy (Davis et al., 1971).

Figure 2 shows a field containing two heteroduplexes (identified by arrows) together with full length single- and double-stranded circular *WT* markers. Each of the heteroduplexes has one long and one very short duplex segment (A and D, respectively) and two single-stranded

TABLE 1

Relative quantities and sizes of defective mutant DNAs obtained following co-infections with *tsA30* virus helper at 41°

Mutants from <i>EcoRI</i> endonuclease-resistant pool	Percent of total DNA as defective mutant*	SV40 fractional length (mean \pm standard error) [†]
<i>dl</i> -811 [‡]	52	0.706 \pm 0.004; 0.861 \pm 0.004 ^δ
<i>dl</i> -812	85	0.886 \pm 0.003
<i>dl</i> -813	85	0.866 \pm 0.003
<i>dl</i> -816	90	0.923 \pm 0.003
<i>dl</i> -817	87	0.923 \pm 0.007
Mutants from <i>HpaII</i> endonuclease-resistant pool		
<i>dl</i> -814	87	0.889 \pm 0.004
<i>dl</i> -815	83	0.956 \pm 0.003
<i>dl</i> -818	96	0.820 \pm 0.006
<i>dl</i> -819	95	0.882 \pm 0.004
<i>dl</i> -820	94	0.926 \pm 0.004
<i>dl</i> -821	79	0.888 \pm 0.008

* These values represent the proportion of circular DNA obtained from each mixed infection that was resistant to cleavage by either the *EcoRI* or *HpaII* restriction endonuclease as determined by sedimentation in neutral CsCl gradients.

[†] The standard deviations were all less than 4% of their means.

[‡] The cloned defective mutants have been numbered 811-821, with *dl* = deletion and - meaning complementation group not yet determined, following the nomenclature suggested by Robb et al., 1972.

^δ This mutant yielded two populations of *EcoRI*-resistant DNA. Both have the same sequence deleted, but the smaller one lacks the duplicated segment found in the other (Mertz and Berg, unpublished).

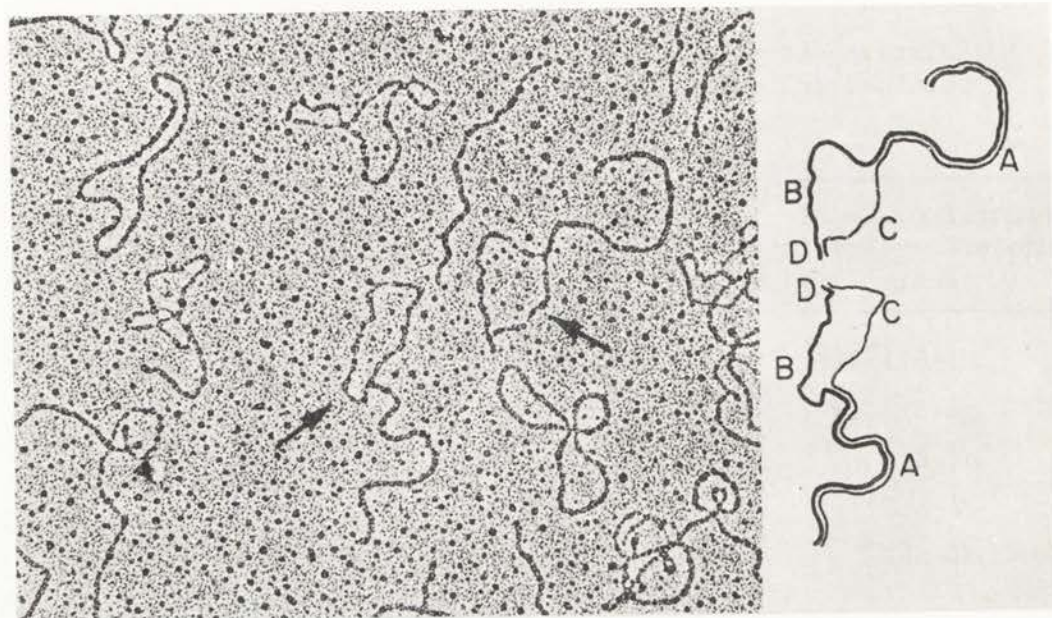


Figure 2. Two-stranded heteroduplexes formed after denaturation and renaturation of *Hpa*II endonuclease cleaved mutant *dl*-812 and *WT* DNAs. The bold and lighter lines of the drawing represent the *WT* and mutant strands, respectively. See text for further description.

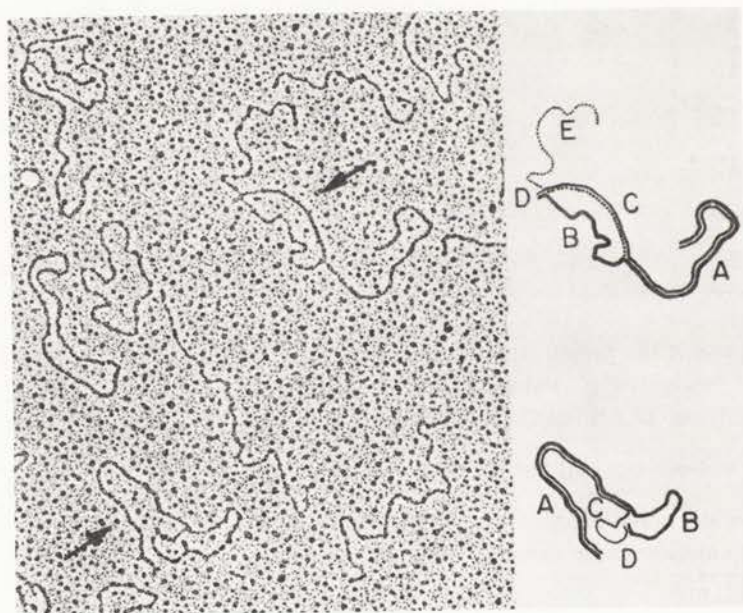


Figure 3. A three-stranded heteroduplex corresponding to the molecules in Figure 2. The dotted line in the drawing represents the third strand hybridized to the duplicated segment.

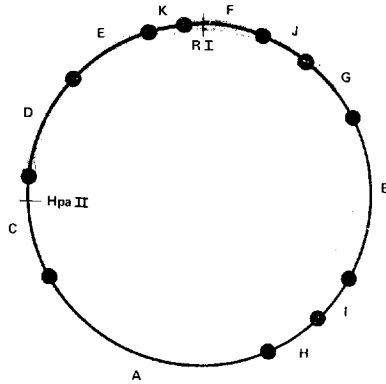


Figure 4. Extent of deleted and duplicated segments in mutant *dl-812*. The darker shaded portion covers the deleted segment, the lighter shaded region denotes the duplicated segment, the solid circles identify the *Hind* II + III cleavage sites (Danna and Nathans, 1971), and the *Hpa*II and *Eco*RI cleavage sites are shown by lines.

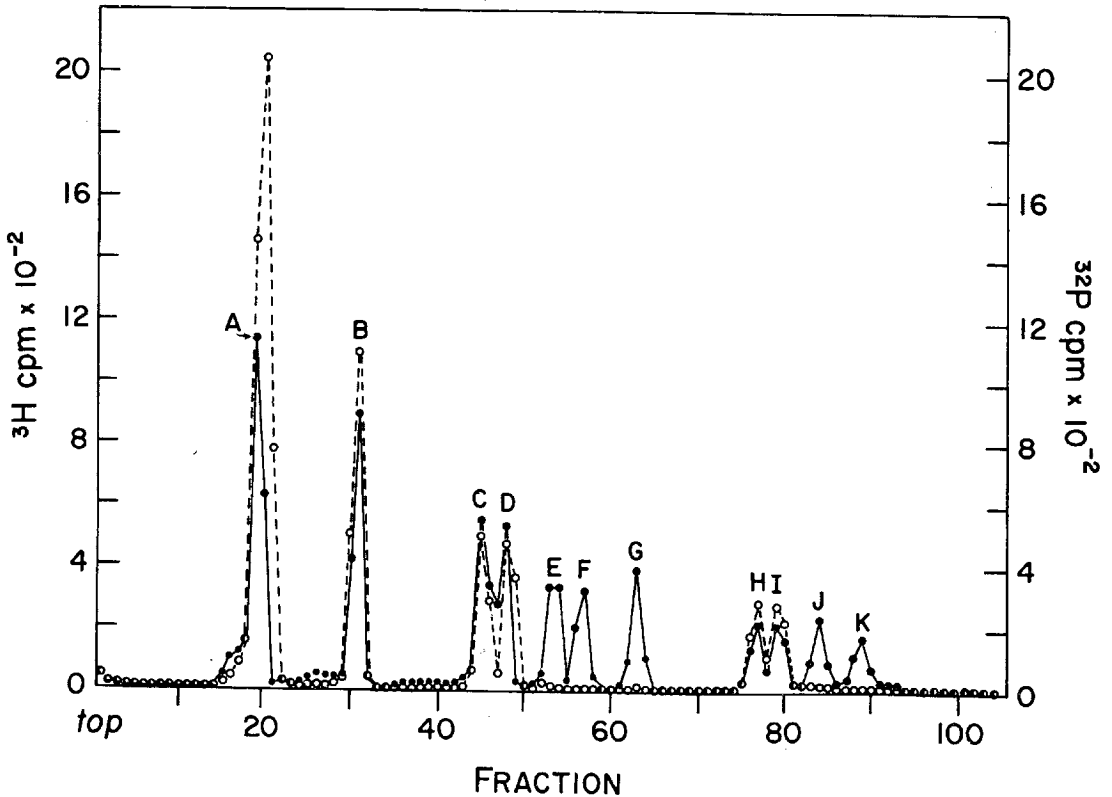


Figure 5. Polyacrylamide gel electrophoresis patterns of the *Hind* II + III fragments produced from a mixture of ³²P-WT DNA (● — ●) and ³H-mutant *dl-812* DNA (○ — ○).

segments of unequal length (B and C) that form a substitution loop. Because the sum of the lengths of segments A, D and B equals the wild-type length, segment B must be the sequence of the wild-type strand that is deleted from the mutant strand. Since the ends of the heteroduplexes are derived from the *Hpa*II restriction sequence (map position 0.735) and the deletion covers the *Eco*RI restriction site, we may fix the coordinates of the deleted portion as 0.76 to 0.15, or a loss of 0.39 SV40 fractional length.

What is the other single-stranded segment (C) of the substitution loop? The occurrence of another type of heteroduplex (Figure 3) provides the answer. In this structure segment C has formed a duplex with a third SV40 strand, showing that segment C is viral in origin. Therefore, mutant *dl*-812 contains a duplication as well as a deletion. From measurements of the lengths of the single-stranded (E) and duplexed (C) regions of the third strand, we can deduce that segment C of the mutant strand corresponds to the wild-type sequence defined by the coordinates 0.45 to 0.72. (Because the mutant does not contain two *Hpa*II restriction sites, the duplication cannot extend beyond the duplexed portion of segment C, but it probably terminates very close to 0.735.)

Figure 4 summarizes the estimated extents of the deletion (the heavily shaded segment) and of the duplication (the lightly shaded segment) as well as the cleavage sites for *Hind* II + III endonuclease (Danna and Nathans, 1971). A *Hind* II + III digest of mutant *dl*-812 DNA should lack fragments D, E, K, F and J, and fuse approximately half of fragment G to the portion of the C fragment that was duplicated; moreover, an extra fragment almost as large as the A fragment should be present. The polyacrylamide gel electrophoresis fragment pattern obtained from a *Hind* II + III digest of a mixture of ³H-labeled mutant *dl*-812 and ³²P-labeled *WT* DNA (Figure 5) reveals that fragments E, K, F, J and G are not present in the mutant pattern and that there is a new fragment migrating slightly faster than the wild-type A band; furthermore, a normal C fragment is present and another fragment occurs in place of the missing D band. Our estimate of the coordinates for the duplication predicts a modified C fragment (0.65 to 0.72, or 0.07 SV40 fractional length) fused to the remaining portion of the G fragment (0.15 to 0.17, or 0.02 SV40 fractional length) to yield a fragment having a mobility slightly less than the D fragment (0.095 SV40 fractional length). Thus, the *Hind* II + III fragments produced with mutant *dl*-812 DNA appear to be those predicted for molecules having the structure proposed in Figures 4 and 6.

We may speculate that mutant *dl*-812 arose during propagation of the virus at high multiplicity from a tandem duplication of the

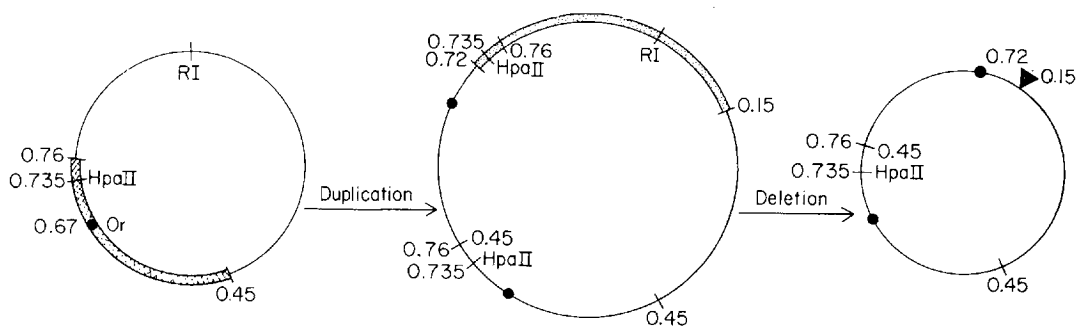


Figure 6. Proposed sequence of duplication and deletion events resulting in formation of mutant *dl-812*. The filled circle on the SV40 map represents the position of the replication origin deduced by Danna and Nathans (1972) and Fareed et al. (1972). The cross-hatched segment indicates the extent of the initial duplication and the stippled segment defines the extent of the subsequent deletion (indicated in the final circle by the solid triangle with coordinates attached to it).

region 0.45 to 0.76 followed by a deletion beginning in the duplication (at 0.72) and extending to 0.15 (Figure 6). This leaves a molecule with one *HpaII* restriction sequence, no *EcoRI* restriction site, a duplicated segment extending from 0.45 to 0.72 and a deletion from 0.76 to 0.15, for a total 0.88 SV40 fractional length. It is clearly defective in that it lacks the region from 0.76 to 0.15, a portion of the late region now believed to contain the D and B complementation groups (Lai and Nathans, personal communication; Shenk, Rhodes, Rigby and Berg, unpublished).

Individual clones of viable defective SV40 virus

Each of the cloned *HpaII* endonuclease-resistant mutants referred to above fails to produce plaques in the absence of helper virus, but plaque-forming DNA molecules can be isolated from the *HpaII* endonuclease-resistant pool of defective DNA which served as the starting material. Plaques produced from these are distinguishable from *WT* plaques by their smaller size, their late appearance and their slower rate of development.

After serial plaque-purification, 10 individual viral clones repeatedly displayed the small plaque phenotype (Figure 7); as anticipated, the viral DNA was completely resistant to cleavage by *HpaII* restriction endonuclease under conditions where more than 95% of the *WT* DNA is cleaved.

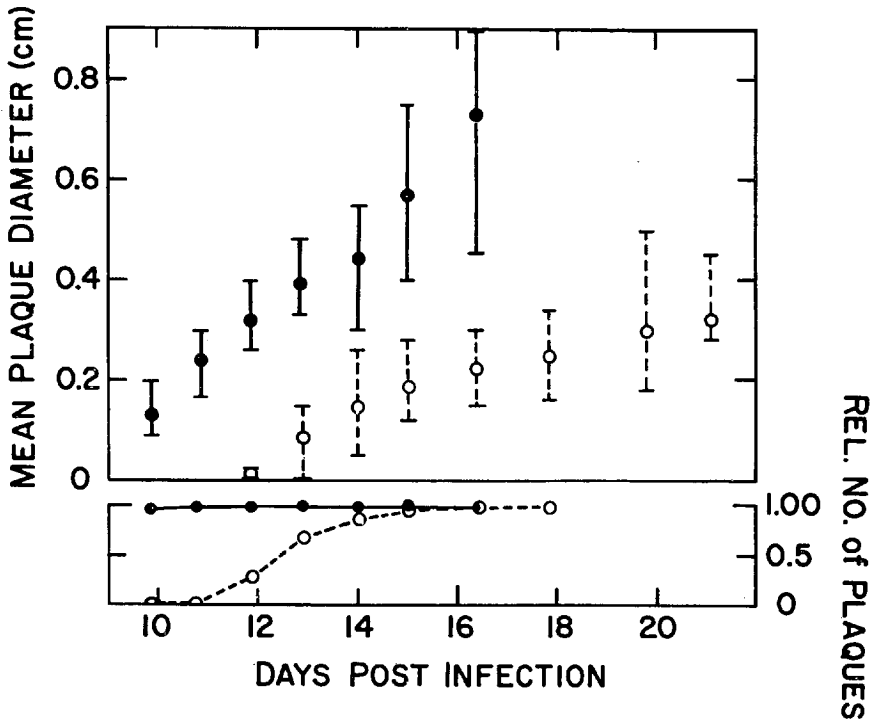


Figure 7. Size and rate of development of mutant and wild-type virus plaques. Top graph: the mean plaque diameter (and range) of 10 wild-type (●) and mutant (○) plaques as a function of time after infection. Bottom graph: the number of wild-type (●) and mutant (○) plaques at a particular time relative to the number of plaques observed 16 days after infection.

Is the resistance to *Hpa*II endonuclease cleavage due to a base change in the restriction sequence or to a small deletion which removed the restriction site? This question was answered by comparing the electrophoretic mobility of mutant and *WT* fragments produced by *Hind* II + III cleavage of the respective DNAs. According to Danna and Nathans (1971) the *Hpa*II restriction sequence is contained within the *Hind* C fragment generated by *Hind* II + III cleavage at the 0.65 and 0.76 map positions (see Figure 3). Figure 8 illustrates the polyacrylamide gel electrophoresis fragment patterns obtained from a *Hind* II + III digest of a mixture of two representative ^3H -labeled mutant DNAs and ^{32}P -labeled *WT* DNA. In each case, the mutant DNA digest lacks the *Hind* C fragment and contains a new, faster migrating fragment (C'). Since electrophoretic mobility is inversely proportional to molecular length (Figure 9), we conclude that the *Hind* C' fragments produced with the 10 mutant DNAs are between 80 (the smallest deletion) to 190 (the largest deletion) base pairs shorter than the *WT* *Hind* C fragment. Since the *Hind* D fragment is

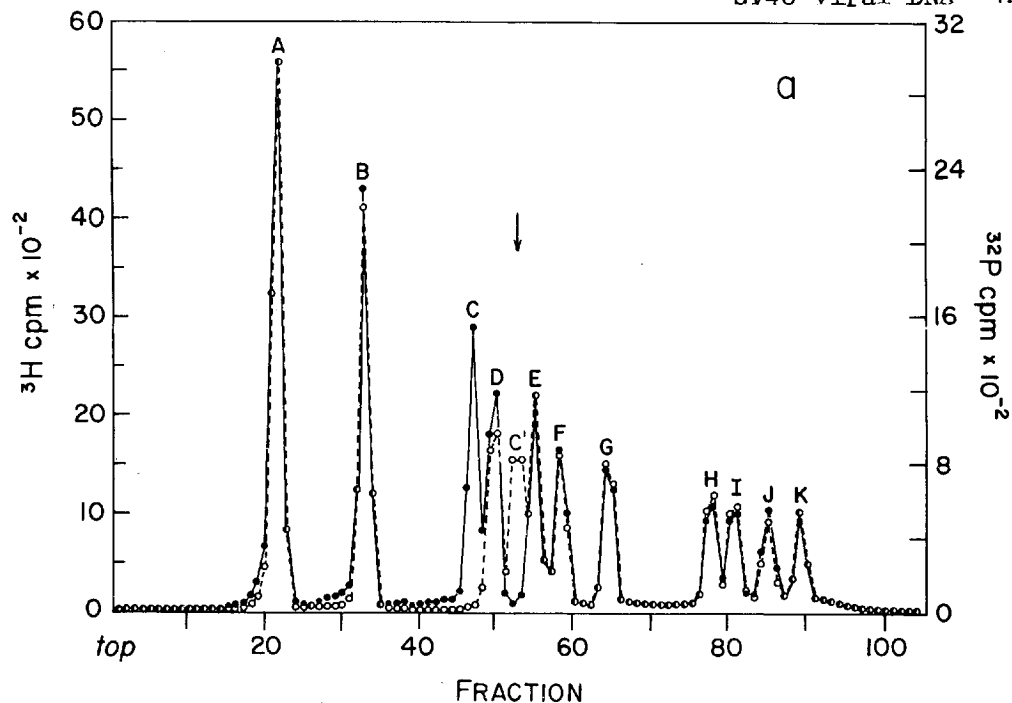
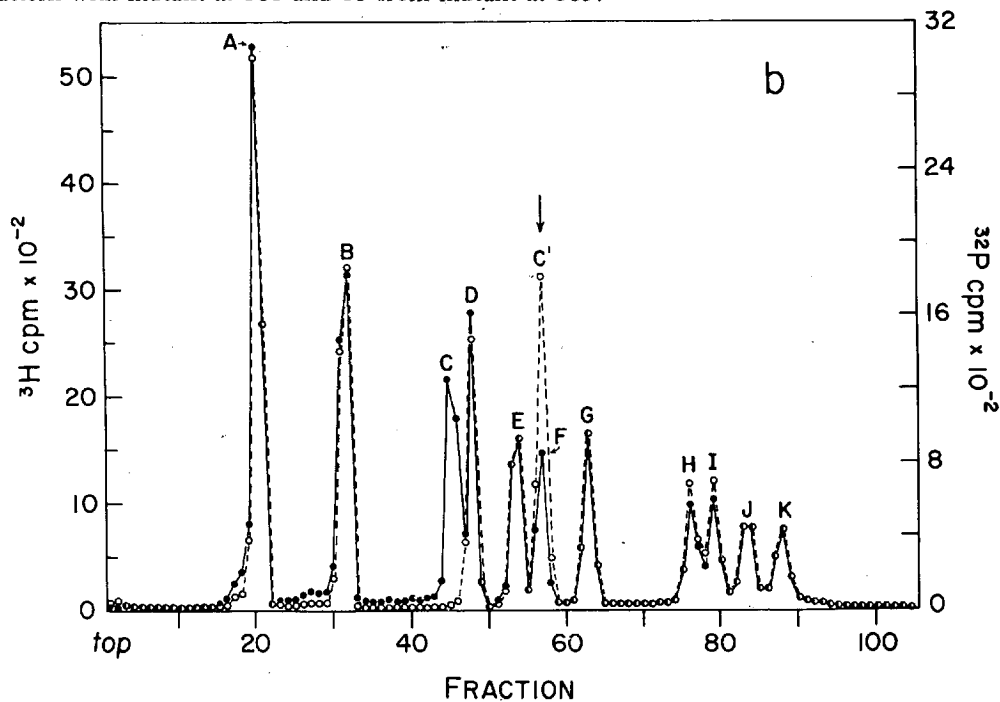


Figure 8. Polyacrylamide gel electrophoresis patterns of the *Hind* II + III fragments produced from a mixture of ^{32}P -WT (●—●) and ^3H -mutant DNA (○—○). 8a is the pattern with mutant *dl*-801 and 8b from mutant *dl*-805.



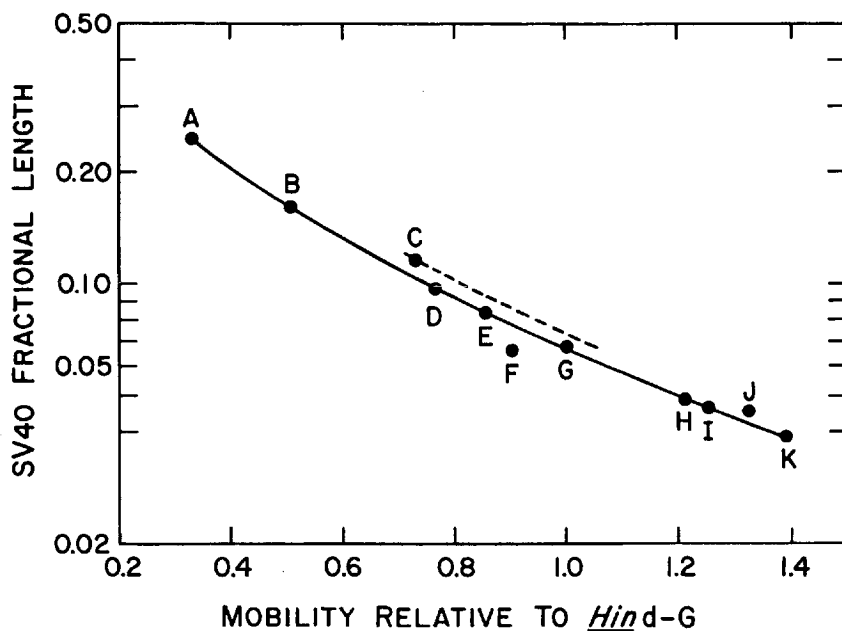


Figure 9. Correlation of electrophoretic mobility and molecular lengths of *Hind* II + III fragments of SV40 DNA. The mobility of each fragment is expressed relative to *Hind* fragment G. The molecular lengths were estimated from the fraction of the total DNA present in each peak from at least ten different electrophoretic analyses with the ^{32}P -labeled WT DNA.

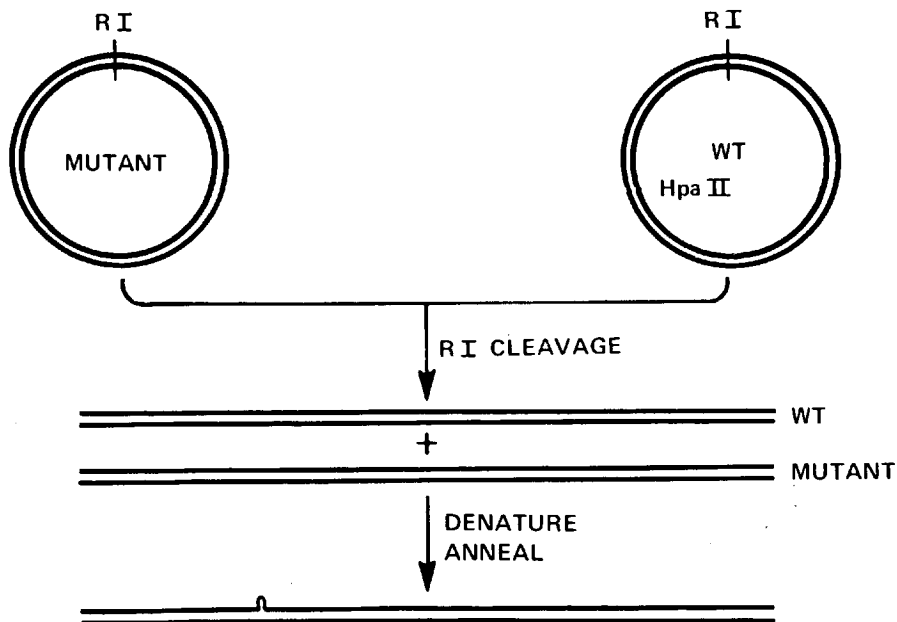


Figure 10. Mapping of deletions of the *Hpa*II restriction site of SV40 DNA.

unaffected in any of the mutants, the *Hind* site at 0.75 must be intact.

The larger deletions can also be detected by heteroduplex analysis in the electron microscope. Denaturation and renaturation of a mixture of *Eco*RI-cut mutant and *WT* DNAs would be expected to yield linear heteroduplexes containing a small deletion loop at the region of the *Hpa*II restriction site (Figure 10). This expectation was confirmed by electron microscopic analysis of heteroduplexes formed with the mutant having the largest deletion (Figure 11); approximately 40% of the linear duplexes contained a barely visible discontinuity at about 0.245 SV40 fractional length from the nearer end (Figure 12).

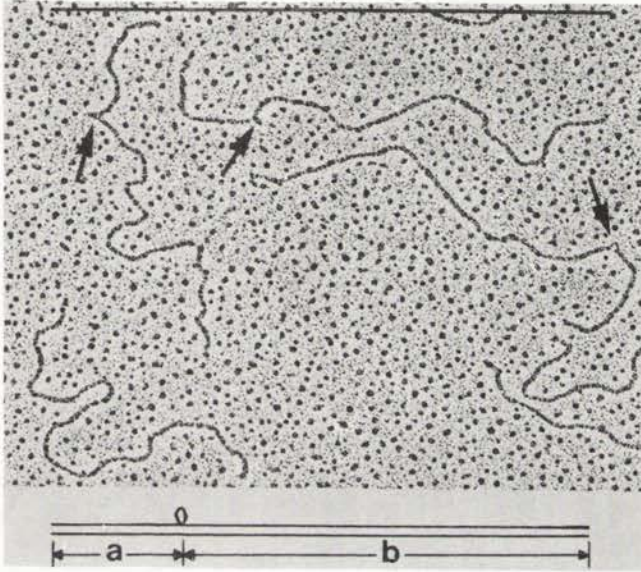


Figure 11. Electron micrograph of heteroduplexes formed as illustrated in Figure 10. The arrows point to the putative deletion loops.

Based on the location of the deletion loop (Figure 12) and the estimate of the size of the deletion from the shortening of the *Hind* C fragment, the map coordinates of the deletions can be deduced (Figure 13). Mutant *dl*-810 is lacking the segment of DNA from 0.720 to 0.755. This assignment is consistent with the fact that the *Hind* II + III and *Hpa*I-restriction sites at 0.760 are still present in all the mutants.

At present we cannot comment on the physiological defect caused by these small deletions. Further studies to clarify this point are in progress.

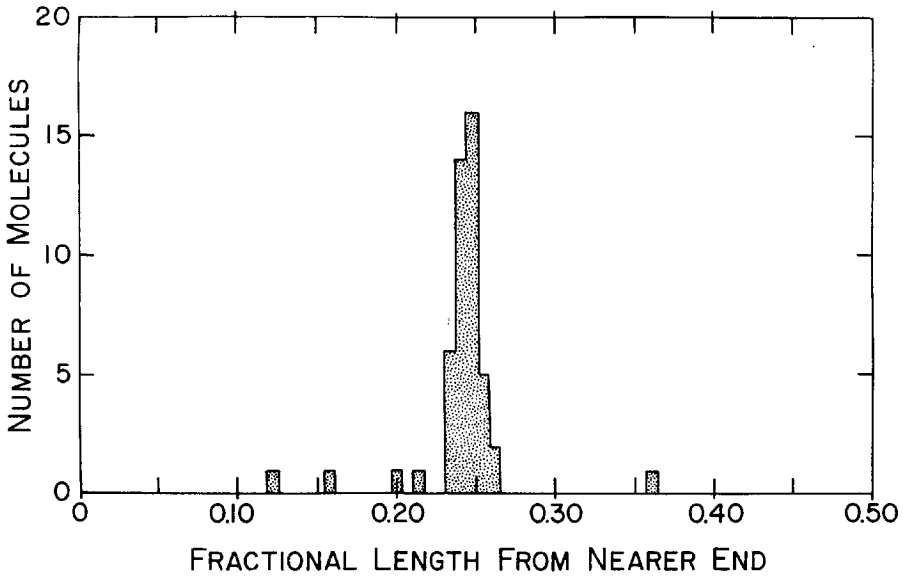


Figure 12. Location of the deletion loop in heteroduplexes between *EcoRI* endonuclease-cut DNA from mutant *dl-810* and *WT SV40* DNA. The histogram presents the number of molecules having a visible discontinuity along the length against their distance from the nearer end (segment a in Figure 11).

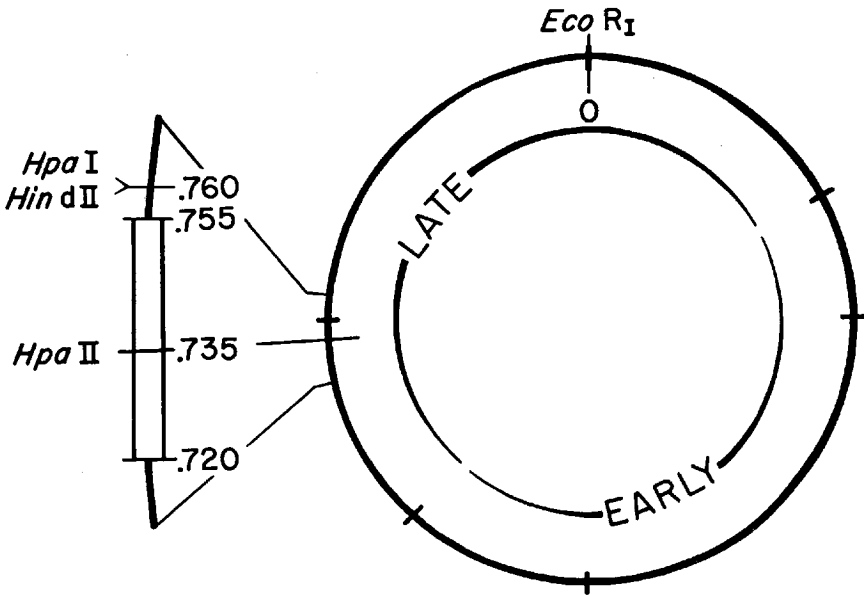


Figure 13. SV40 DNA map location of a deletion (*dl-810*) which removes the *HpaII* restriction site and causes altered plaque morphology.

Construction of insertion mutants *in vitro*

We have developed a general procedure for constructing mutants of SV40 virus by covalently inserting a short segment of extraneous DNA into the viral DNA molecule at specific locations. The supposition is that such an insert interrupts the normal sequence of nucleotides and either partially or completely inactivates the genetic function of that region. Moreover, the map position of such inserts should be readily discernible by heteroduplex analysis.

The overall protocol for constructing such molecules *in vitro* is a modification of the earlier procedure of Jackson et al. (1972). In this procedure, however, the homopolymer chains themselves serve as the insert (Figure 14).

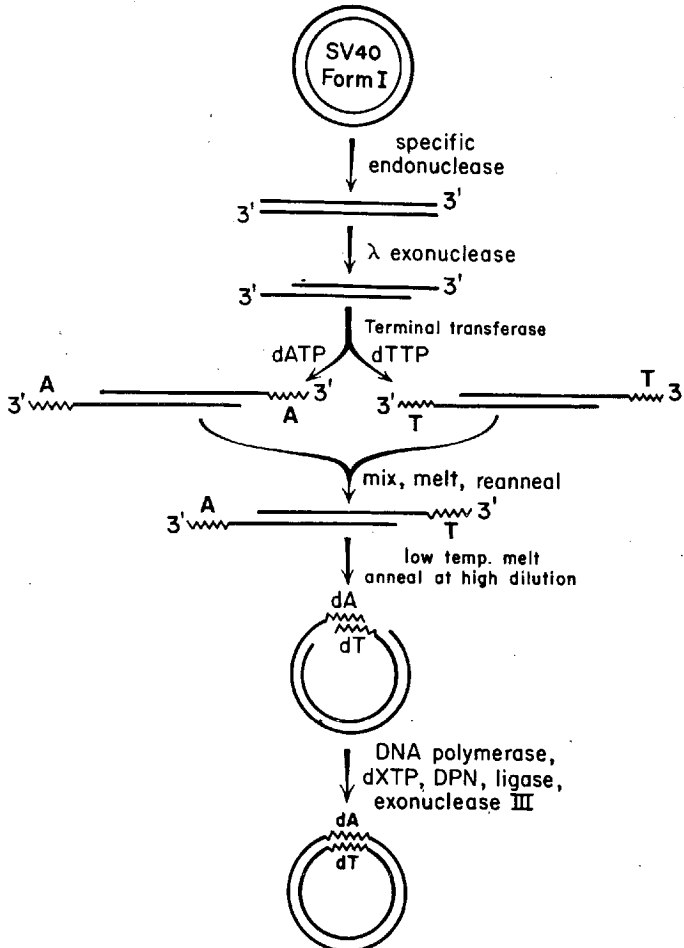


Figure 14. Reaction scheme for the synthesis of covalently closed circles of SV40 DNA containing an insert of poly(dA:dT).

Briefly, the procedure involves the following steps: a) opening of the SV40 circular DNA by a double-strand scission at a single location (in the example to be discussed here, by the *Hpa*II restriction endonuclease); b) treatment with λ phage 5'-exonuclease to remove about 25 nucleotides from each 5' terminus; c) reaction of the product with either dATP or dTTP and the purified terminal transferase enzyme to attach short (about 50 residues) homopolymer 'tails' of poly(dA) or poly(dT) to the 3' termini; d) denaturation and annealing of an equimolar mixture of poly(dA)- and poly(dT)-ended duplex molecules to generate molecules containing a poly(dA)-tail at one 3'-end and a poly(dT)-tail at the other 3'-end of the same duplex; e) cyclization of these linear molecules by their newly created cohesive ends and covalent closure of the hydrogen-bonded circular molecules mediated by DNA polymerase, the four deoxynucleoside triphosphates and DNA ligase. By limiting the extent of addition of poly(dA) and poly(dT) 'tails' to about 50 nucleotides, the final product is still small enough to permit encapsidation of the modified DNA molecule and therefore its propagation in virions.

Not unexpectedly, insertion of the short poly(dA:dT) sequence at the *Hpa*II restriction site renders the circular molecule resistant to cleavage by *Hpa*II endonuclease. Although our proof that there is an insert of poly(dA:dT) at 0.735 map position is not complete, heteroduplex analysis strongly supports this contention. The principle for mapping the poly(dA:dT) insertion is similar to that used to detect the small deletions at the *Hpa*II restriction site (see Figure 10). The molecules containing the poly(dA:dT) inserts can be cut to linear molecules with *Eco*RI endonuclease and heteroduplexed to *Eco*RI-cut *WT* DNA. Such heteroduplexes should contain a small loop of single-stranded poly(dA) and others a loop of poly(dT) (of course some of the reannealed products are homoduplexes of two *WT* or two modified strands). However, the small size of the insertion loops makes their direct visualization extremely unlikely. The poly(dA) loops could be visualized, however, by annealing linear SV40 molecules having poly(dT) tails (those shown in Figure 14) to the heteroduplexes. This yields branched molecules (Figure 15) in which the unit length homoduplex is attached by poly(dA:dT) base pairs to the poly(dA) insertion loop of the heteroduplex. Measurements of the three arms (Figure 15) of the branched molecule indicate that a unit length SV40 linear DNA molecule's poly(dT) tail can anneal to the heteroduplex at the position of the poly(dA:dT) insertion (0.735 map position).

Molecules containing the poly(dA:dT) insert are infectious and produce plaques in the absence of the helper virus. Moreover, the poly(dA:dT) insert remains in the viral DNA after repeated infections.

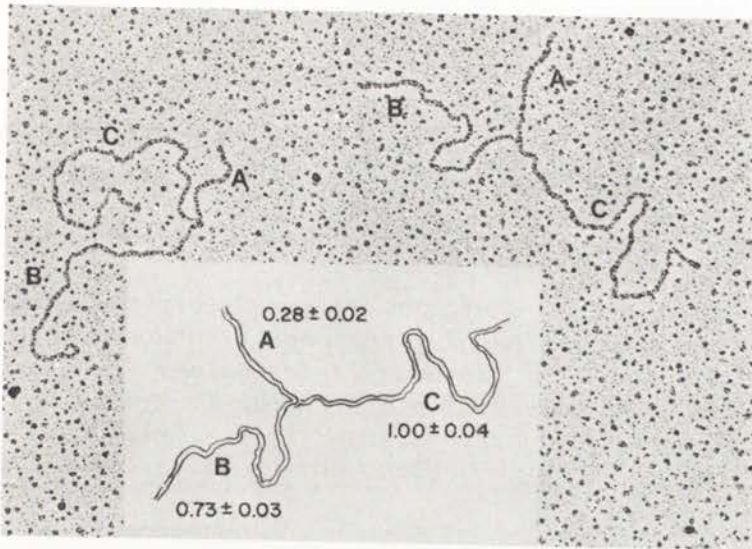


Figure 15. Electron microscopic visualization of small poly(dA) insertion loops present in heteroduplexes. Segments A and B make up the linear heteroduplex formed from the *Eco*RI endonuclease-cut *WT* DNA and the *Eco*RI-cut DNA containing the poly(dA:dT) insert at the *Hpa*II restriction site (0.735 map position). Segment C is the linear SV40 DNA molecule bound to the poly(dA) loop by its poly(dT) tail.

The plaque morphology and rate of plaque development is virtually indistinguishable from those produced by the naturally occurring variants which lack this region. Since neither insertion of a new sequence, nor deletion of an existing sequence at map position 0.735 prevents viral multiplication, this region is either part of a regulatory sequence or it codes for an unessential portion of one of the late proteins.

Additional experiments designed to introduce poly(dA:dT) segments at other locations are in progress. The analysis of the phenotypes of such insertion mutants together with the studies of the naturally occurring defectives should aid in establishing a complete physico-genetic map of SV40 DNA.

ACKNOWLEDGMENTS

This work was supported by research grants from the National Institutes of Health (GM 13235-09) and from the American Cancer Society (VC 23C). J.E.M. is a trainee of the National Institute of Health (5 TI GM 196-15). J.C. is a member of the Department of Biological Sciences, University of California, Santa Barbara, California, and was supported in part by NSF 23429.

REFERENCES

- Chou, J.Y. and R.G. Martin. 1974. Complementation analysis of SV40 mutants. *J. Virol.* In press.
- Danna, K. and D. Nathans. 1971. Specific cleavage of SV40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc. Nat. Acad. Sci. USA* 68: 2913.
- , ———. 1972. Bidirectional replication of Simian Virus 40 DNA. *Proc. Nat. Acad. Sci. USA* 69: 3087.
- Davis, R.W., M. Simon and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *In Methods in Enzymology* (L. Grossman and K. Moldave, ed.) Vol. 21D, p.413. Academic Press, New York.
- Fareed, G.C., C.F. Garon and N.P. Salzman. 1972. Origin and direction of Simian Virus 40 deoxyribonucleic acid replication. *J. Virol.* 10: 484.
- Girardi, A.J. 1965. Prevention of SV40 virus oncogenesis in hamsters. I. Tumor resistance induced by human cells transformed by SV40. *Proc. Nat. Acad. Sci. USA* 54: 445.
- Hedgpeth, J., H.M. Goodman and H.W.P. Boyer. 1972. DNA nucleotide sequence restricted by the RI endonuclease. *Proc. Nat. Acad. Sci. USA* 69: 3448.
- Jackson, D.A., R.H. Symons and P. Berg. 1972. A biochemical method for inserting new genetic information into SV40 DNA: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *E. coli*. *Proc. Nat. Acad. Sci. USA* 69: 2904.
- Khoury, G., M.A. Martin, T.N.H. Lee, K.J. Danna and D. Nathans. 1973. A map of Simian Virus 40 transcription sites expressed in productively infected cells. *J. Mol. Biol.* 78: 377.
- Kimura, G. and R. Dulbecco. 1972. Isolation and characterization of temperature-sensitive mutants of Simian Virus 40. *Virology* 49: 394.
- Kit, S., S. Tokuno, K. Nakajima, D. Trkula and D.R. Dubbs. 1970. Temperature-sensitive Simian Virus 40 mutant defective in late function. *J. Virol.* 6: 286.
- Morrow, J.F. and P. Berg. 1972. Cleavage of Simian Virus 40 DNA at a unique site by a bacterial restriction enzyme. *Proc. Nat. Acad. Sci. USA* 69: 3365.
- , ———. 1973. The location of the T4 gene 32 protein binding site on Simian Virus 40 DNA. *J. Virol.* 12: 1631.
- Morrow, J.F., P. Berg, T.H. Kelly, Jr and A.M. Lewis, Jr. 1973. Mapping of Simian Virus 40 early functions on the viral chromosome. *J. Virol.* 12: 653.
- Mulder, C. and H. Delius. 1972. Specificity of the break produced by restricting endonuclease RI in SV40 DNA as revealed by partial denaturation. *Proc. Nat. Acad. Sci. USA* 69: 3215.

- Nathans, D. and K.J. Danna. 1972. Specific origin of SV40 DNA replication. *Nature New Biol.* 236: 200.
- Radloff, R., W. Bauer and J. Vinograd. 1967. A dye buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc. Nat. Acad. Sci. USA* 57: 1514.
- Robb, J.A. and R.G. Martin. 1970. Genetic analysis of Simian Virus 40. I. Description of microtitration and replica-plating techniques for virus. *Virology* 41: 751.
- , ———. 1972. Genetic analysis of Simian Virus 40. III. Characterization of a temperature-sensitive mutant blocked at an early stage of productive infection in monkey cells. *J. Virol.* 9: 956.
- Robb, J.A., H.S. Smith and C.D. Scher. 1972. Genetic analysis of Simian Virus 40. IV. Inhibited transformation of BALB 3T3 cells by a temperature-sensitive early mutant. *J. Virol.* 9: 969.
- Sack, G.H., Jr and D. Nathans. 1973. Cleavage of SV40 DNA by restriction endonuclease from *Haemophilus parainfluenzae*. *Virology* 51: 517.
- Sambrook, J., B. Sugden, W. Keller and P.A. Sharp. 1973. Transcription of Simian Virus 40. III. Mapping of 'early' and 'late' species of RNA. *Proc. Nat. Acad. Sci. USA* 70: 3711.
- Sharp, P.A., B. Sugden and J. Sambrook. 1973. Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* 12: 3055.
- Smith, H.O. and D. Nathans. 1973. A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J. Mol. Biol.* 81: 419.
- Subramanian, K.N., R. Dhar, J. Pan and S. Weissman. 1974. The mapping and ordering of fragments of SV40 DNA produced by restriction endonucleases. In *Procedures in Nucleic Acids Research* (G.L. Cantoni and D.R. Davies, ed.) Harper and Row, New York. In press.
- Tai, H.T., C.A. Smith, P.A. Sharp and J. Vinograd. 1972. Sequence heterogeneity in closed SV40 DNA. *J. Virol.* 9: 317.
- Tegtmeyer, P. 1972. SV40 DNA synthesis: The viral replicon. *J. Virol.* 10: 591.
- Tegtmeyer, P. and H.L. Ozer. 1971. Temperature-sensitive mutants of SV40: Infection of permissive cells. *J. Virol.* 8: 516.
- Yoshiike, K. 1968. Studies on DNA from low-density particles of SV40. I. Heterogeneous defective virions produced by successive undiluted passages. *Virology* 34: 391.
- Yoshiike, K. and A. Furuno. 1969. Heterogeneous DNA in Simian Virus 40. *Fed. Proc.* 28: 1899.
- Yoshimori, R.N. 1971. A genetic and biochemical analysis of the restriction and modification of DNA by resistance transfer factors. Ph.D. dissertation, University of California, San Francisco Medical Center.

Mapping of mutational alterations in DNA with S1 nuclease: the location of deletions, insertions and temperature-sensitive mutations in SV40

Thomas E. Shenk, Carl Rhodes, Peter W. J. Rigby and Paul Berg

S1 nuclease, a single-strand specific nuclease from *Aspergillus oryzae*, can be used to map accurately mutational changes in SV40 which are too small to detect by conventional electron microscopic analysis of heteroduplex DNAs. In order to map a mutational change, unit length, linear DNA was prepared from both the SV40 mutant and its wild-type parent. After mixing, the DNAs were denatured and reannealed to form heteroduplexes. S1 nuclease can then recognize and cut at the site of non-base-paired regions in the heteroduplexes. Heteroduplex DNA molecules containing a small deletion loop (about 190 bases) or insertion loop (about 50 bases) were cleaved by S1 nuclease at the site of the single-stranded loops, producing two fragments whose lengths reflected the position of the deletion or insertion. Furthermore, S1 nuclease could cleave a heteroduplex formed between the DNAs of an SV40 temperature-sensitive mutant and either its revertant or wild-type parent. This activity of S1 nuclease has been exploited to map the positions of the SV40 *tsA30* and *tsD202* mutations.

S1 nuclease from *Aspergillus oryzae* degrades single- but not double-stranded DNA (Ando, 1966); nevertheless, superhelical SV40 (Beard et al., 1973) and polyoma (Germond et al., 1974) DNA are converted to unit length linear molecules by this enzyme. Presumably, this occurs because unpaired, or weakly hydrogen-bonded regions susceptible to S1 nuclease, occur or can be induced in the strained superhelical molecule. S1 nuclease can also cleave duplex DNA at or near single-strand nicks under conditions where intact duplex DNA is relatively resistant (Beard et al., 1973; Germond et al., 1974).

These properties prompted us to test whether S1 nuclease could be used to map the location of small deletions, insertions, or, in fact, any difference in base sequence between otherwise homologous DNAs. If S1 nuclease can cleave at non-base-paired regions in DNA molecules, heteroduplex DNA molecules containing mismatched regions should also be cleaved. For example, heteroduplex DNA molecules formed from the complementary strands of a deletion mutant and wild-type SV40 DNA would contain a single-strand loop at a position corresponding to the deletion (Figure 1). S1 nuclease digestion of

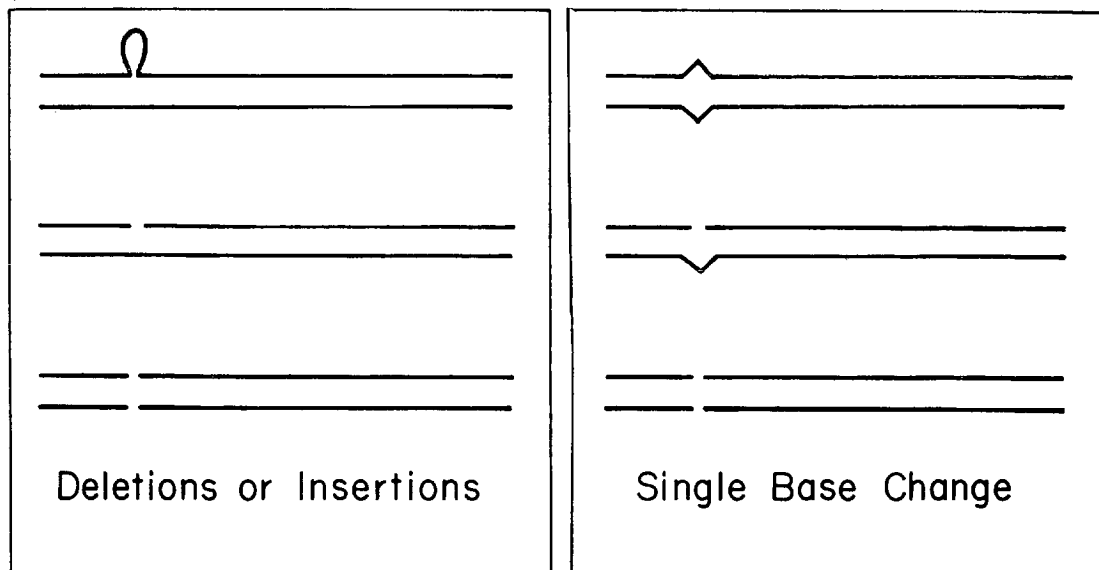


Figure 1. Heteroduplexes which contain mismatched regions susceptible to cleavage by S1 nuclease. A duplex DNA with a single nick is proposed as an intermediate in the S1 nuclease cutting reaction. There is as yet no direct evidence for such an intermediate.

this single-strand loop and subsequent cleavage of the intact strand at or near the nick (or gap) should yield fragments whose length would reflect the position of the deletion loop. Insertion mutations could be similarly located. The experiments we report here show that these expectations are fulfilled; S1 nuclease has been used to locate a deletion of about 190 base pairs and an insertion of about 50 base pairs. Of greater interest is the finding that S1 nuclease can also cleave a heteroduplex formed from the DNAs of an SV40 temperature-sensitive mutant and either its revertant or wild-type parent. The map positions we have determined for the SV40 *tsA30* and *tsD202* mutations by this method are consistent with the locations found by other means (Lai and Nathans, 1975). Although we cannot specify how the base sequence of the temperature-sensitive and revertant DNAs differ from each other or from wild-type DNA (are they only a single base change? Figure 1), we are optimistic that this method will be useful for mapping accurately point mutations, making it unnecessary to rely on genetic recombination experiments.

MATERIALS AND METHODS

Cells and viruses The origin and the procedures for the growth of primary African Green Monkey kidney cells (AGMK) and the established

monkey kidney cell lines (CV-1P and MA-134) have been described (Mertz and Berg, 1974a). All virus stocks and virus DNA were prepared in MA-134 cells. Plaque assays were performed on CV-1P cells. Mutant *dL-808* is a spontaneous deletion mutant isolated from a stock of SV40 strain Rh911 which had been serially passaged without dilution through AGMK cells (Mertz and Berg, 1974b; Mertz et al., 1975). SV40 mutants *tsA30* and *tsD202* as well as their wild-type parents were obtained from Peter Tegtmeyer (Tegtmeyer, 1972) and Robert Martin (Chou and Martin, 1974) respectively. Temperature-sensitive mutants were grown at 32°C. Spontaneous revertants of these mutants were from plaques that arose on cell monolayers incubated at 41°C.

Enzymes S1 nuclease was prepared from 'Enzopharm' powder (Enzyme Development Corp., N.Y.) by the procedure of Vogt (1973). One unit of S1 nuclease releases 1.0 nmole of nucleotides per minute at 37°C when acting on sonicated, denatured salmon sperm DNA at pH 4.4. *EcoRI* restriction endonuclease was prepared by the procedure of Greene et al. (1974). *HpaI* and *HpaII* restriction endonucleases were prepared by the procedure of Sharp et al. (1973) with minor modifications.

DNA substrates SV40 DNA was extracted by the procedure of Hirt (1967) from MA-134 cells infected at a multiplicity of <0.05 pfu/cell when >90% of the cells showed cytopathic effect. Form I viral DNA was purified directly from the Hirt supernatant by adding CsCl to 1.56 g/cc and ethidium bromide to 200 µg/ml and centrifuging to equilibrium. The band of Form I DNA was collected, and the ethidium bromide was removed by passing the DNA through Dowex-50 (Radloff et al., 1967). Full length linear DNA or DNA fragments were prepared from purified Form I DNA by treatment with restriction endonucleases using published protocols (*EcoRI*, Mertz and Davis, 1972; *HpaI* and *HpaII*, Sharp et al., 1973), and then the DNA was repurified by velocity sedimentation in a neutral CsCl gradient.

Cleavage of heteroduplex DNA with S1 nuclease Heteroduplexes of *EcoRI* or *HpaII* endonuclease-generated linear SV40 DNAs were prepared by denaturing a mixture containing equal parts of two DNA species (1.5 to 6 µg/ml of each) in NaOH (0.1 N). After 10 min at room temperature, the solution was titrated to pH 7 to 8 with HCl (0.1 N), the Na⁺ concentration was raised to 300 mM, and the DNA was reannealed at 68°C for 3 min. The heteroduplexed DNA was treated with S1 nuclease (135 units/µg DNA) in the presence of Zn⁺⁺ (0.5 mM), Na⁺ (280 mM) and CH₃COO⁻ (30 mM) at pH 4.4. After 10 min at 37°C, the reaction was terminated by raising the pH with 0.05 volume of Tris base (2 M), and by increasing the Na⁺ concentration to 500 mM. To decrease the volume and Na⁺ concentration of the sample prior to gel electrophoresis, the

DNA was precipitated at -20°C after the addition of yeast RNA (20 $\mu\text{g}/\text{ml}$) and two volumes of ethanol.

Gel electrophoresis Agarose gels (1%, 6 mm in diameter, 200 mm long) or composite agarose-acrylamide gels (0.7% agarose, 2.0% acrylamide, 0.1% N, N'-methylene-bis-acrylamide) were prepared in Tris-borate buffer (110 mM Tris-OH, 89 mM boric acid, 2.5 mM EDTA, pH 8.2) (Greene et al., 1974; Peacock and Dingman, 1968). Samples were applied in 40 μl of Tris-borate buffer containing sucrose (20% w/v). Electrophoresis was at 40 volts for 16 hours. The DNA bands were stained with ethidium bromide and then visualized using a short wavelength UV light. The fluorescent bands were photographed using a Vivitar orange (02) filter and Kodak TX0135 film. The negatives were then scanned using a Joyce, Loebel and Co. microdensitometer.

RESULTS

S1 nuclease cleaves heteroduplex DNAs at the site of deletion or insertion loops The DNA of SV40 mutant $d\bar{L}$ -808 is resistant to cleavage by the *Hpa*II restriction endonuclease because it contains a deletion of about 190 base pairs between the map coordinates 0.72 and 0.75 (the map coordinates are expressed in SV40 fractional length measured clockwise from the *Eco*RI endonuclease cleavage site) (Mertz and Berg, 1974b; Mertz et al., 1975). Heteroduplexes formed between *Eco*RI endonuclease-cut linear DNA from the mutant and wild-type contain a small deletion loop 0.25 SV40 fractional length from the nearer end (Mertz and Berg, 1974b; Mertz et al., 1975). Therefore, these heteroduplex molecules provide an excellent model with which to test the ability of S1 nuclease to detect and cleave DNA at this type of mismatch.

Figure 2b shows that S1 nuclease does, in fact, cleave heteroduplexes of the type mentioned above, producing predominantly two fragments, 0.24 and 0.73 SV40 fractional length, relative to the mobility of fragments of known size (Figure 2c). A mixture of the corresponding homoduplexes, formed by denaturation and renaturation of each DNA by itself, did not yield detectable quantities of such fragments after S1 nuclease digestion (Figure 2a). The length of the fragments is slightly underestimated since S1 nuclease appears to nibble the ends of duplex DNA molecules during the course of the reaction. The extent of nibbling is difficult to estimate accurately but is about 50 base pairs, or 0.01 SV40 fractional length. Thus, the length of the smaller fragment locates the position of the deletion loop at 0.24 to 0.25 SV40 fractional length from the nearest *Eco*RI endonuclease-generated end. This is almost precisely the map position determined by electron microscopy (Mertz and Berg, 1974b; Mertz et al., 1975).

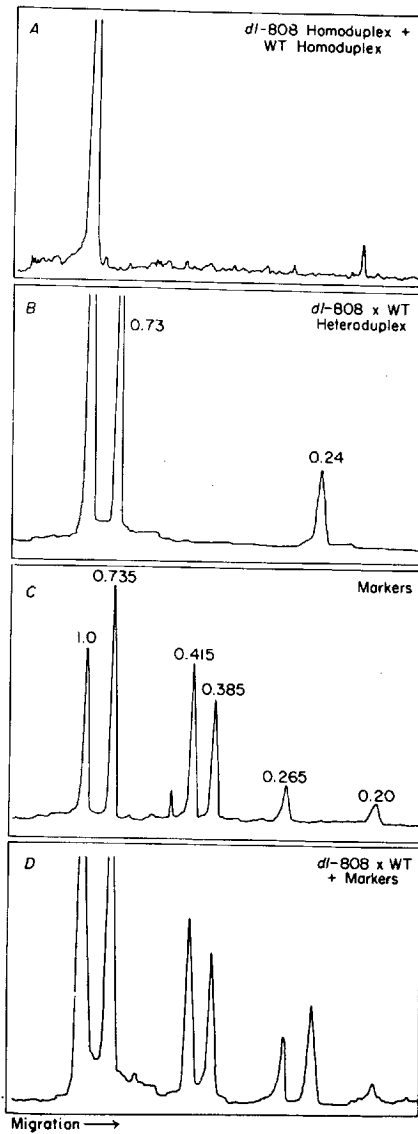


Figure 2. Cleavage by S1 nuclease of heteroduplexes prepared from *EcoRI* endonuclease-generated linear DNAs of *dl-808* and its wild-type parent. The S1 nuclease reactions contained 5 $\mu\text{g/ml}$ DNA. Samples of 0.2 μg DNA were applied to each gel. Electrophoresis was for 16 hours at 40 volts in composite agarose-acrylamide gels. (A) S1 nuclease-treated homoduplexes. (B) S1 nuclease-treated heteroduplexes formed from *dl-808* and wild-type DNAs. (C) marker fragments alone. These include *EcoRI* endonuclease-generated SV40 linears, 0.735 and 0.265 SV40 fractional length fragments obtained by sequential cleavage of SV40 Form I DNA with *HpaII* and *EcoRI* endonucleases, and 0.415, 0.385 and 0.20 SV40 fractional length fragments obtained by cleavage of SV40 Form I DNA with *HpaI* endonuclease. (D) same as (B) with marker fragments (0.15 μg) added.

Mutants of SV40 have also been produced by inserting a sequence of about 50 dA:dT base pairs at the *Hpa*II restriction sequence (Carbon and Berg, 1974; Mertz et al., 1975). Though heteroduplex molecules, formed from *Eco*RI endonuclease-cut wild-type and insertion mutant DNAs contain single-stranded loops of either poly(dT) or poly(dA), their small size precluded their visualization in the electron microscope. Nevertheless, S1 nuclease cleaved such heteroduplex molecules, to yield predominantly two fragments, 0.74 and 0.26 SV40 fractional length (data not shown); the respective homoduplexes were not appreciably cleaved. The small fragment produced from the insertion heteroduplex was, as expected, detectably larger (slower migrating in the gel) than the fragment produced from the deletion heteroduplex.

Thus, S1 nuclease appears able to detect and cleave a duplex DNA, preferentially, at small single-stranded loops and thereby permit the accurate mapping of small deletions, insertions and very likely substitutions.

S1 nuclease can also cleave heteroduplexes formed from wild-type, temperature-sensitive and revertant DNAs The success of S1 nuclease in cleaving heteroduplexes at the position of small single-stranded loops prompted the question of whether the enzyme could cleave at mismatches resulting from single base differences. Lacking DNAs with defined single base changes, we elected to examine the susceptibility of heteroduplexes formed from *Eco*RI endonuclease-cleaved temperature-sensitive (*ts*), revertant (-R) and wild-type DNAs. Both *tsA30* and *tsD202* were induced by hydroxylamine and though the nature and number of changes in the base sequence are not known, we surmise that they are probably single base changes. A spontaneous revertant of each *ts* mutant was recovered from plaques that developed at 41°C.

Initially, we examined the S1 nuclease susceptibility of the heteroduplex formed from *Eco*RI endonuclease-cut *tsA30* and *tsA30-R* DNAs as these genomes were assumed to have the fewest number of base pair differences. As expected, the mixed homoduplexes produced from each DNA separately did not yield specific fragments upon S1 nuclease digestion (Figure 3a). But the heteroduplexes of the *ts* and revertant DNAs were clearly cut to produce two easily discernible fragments of 0.68 and 0.32 SV40 fractional length (Figures 3b and d). This places the mismatch between *tsA30* and its revertant at either 0.32 or 0.68 on the SV40 map. A decision between these two alternatives could be made by examining the S1 nuclease cleavage products of the corresponding heteroduplexes formed from *Hpa*II endonuclease-cleaved mutant and revertant DNAs. [*Hpa*II endonuclease cleaves SV40 DNA once at map position 0.735 (Sharp et al., 1973)]. In this instance the putative mismatch would be expected to be in a different

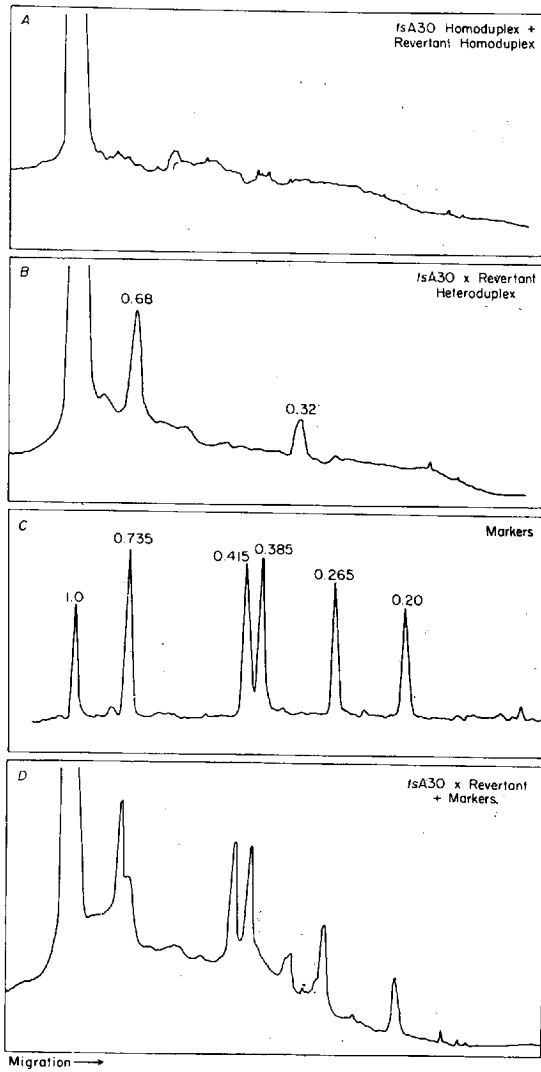


Figure 3. Cleavage by S1 nuclease of heteroduplexes prepared from *Eco*RI endonuclease-generated linear DNAs of *tsA30* and *tsA30-R*. The S1 nuclease reactions contained 11.5 $\mu\text{g/ml}$ DNA. Samples of 0.75 μg DNA were applied to each gel. Electrophoresis was for 16 hours at 40 volts in agarose gels. (A) S1 nuclease-treated homoduplexes. (B) S1 nuclease-treated heteroduplexes formed with *tsA30* and *tsA30-R* DNAs. (C) marker fragments alone. (D) same as (B) with marker fragments (0.15 μg) added.

location relative to the ends of the heteroduplex. Since S1 nuclease cleaved these heteroduplex DNAs to generate fragments of 0.59 and 0.42 SV40 fractional length (data not shown), the mismatch between *tsA30* and its revertant must be located at 0.32 on the SV40 map.

The analysis of the S1 nuclease cleavage products of heteroduplexes between *tsA30* and wild-type DNAs revealed at least two mismatched sequences (Figure 4b). Heteroduplexes formed from *EcoRI* endonuclease-generated linear *tsA30* and wild-type DNAs yielded fragments of 0.62 and 0.38 SV40 fractional length as well as fragments of 0.58 and 0.42 SV40 fractional length following S1 nuclease digestion. None of these fragments were discernible when homoduplexes of these DNAs were reacted with S1 nuclease (Figure 4a). The comparable experiment carried out with *HpaII* endonuclease-generated *tsA30* and wild-type linear DNAs permitted assignment of one of the differences to map position 0.38. This experiment suggests, therefore, that *tsA30* and wild-type DNAs differ in at least two sites and one of these occurs at 0.38 on the SV40 map.

Heteroduplexes, but not the homoduplexes, produced from *tsA30-R* and wild-type DNAs (*EcoRI* endonuclease-cut) also yield specific fragments after cleavage with S1 nuclease (Figure 4c). Preferential cleavage seems to have occurred at both of the sites detected in the heteroduplexes prepared from *tsA30* and wild-type DNAs [0.38 and 0.42 (or possibly 0.58)] and at a third site to yield the same size fragments produced in the heteroduplexes prepared from *tsA30* and *tsA30-R* DNAs, i.e., 0.68 and 0.32 SV40 fractional length. This latter cleavage corresponds to the site at which the *tsA30* and *tsA30-R* sequences differ.

Thus, *tsA30* contains at least two alterations relative to its wild-type parent; these occur at 0.38 and 0.42 (or 0.58) on the SV40 physical map (Figure 6). The alteration causing reversion of the *tsA30* phenotype, therefore, results from a change at a different site, 0.32 map unit.

One inference from these experiments is that the map coordinates 0.3 to 0.4 lie within the A complementation group of SV40 as this is the region within which the mutant and second site revertant differ from the wild-type. This conclusion is reinforced by the findings of Lai and Nathans (1974, 1975) who have determined, by marker rescue experiments, that the *tsA30* mutation is located within *Hind* II + III fragment H (0.375 - 0.43 map position) and the *tsA28* mutation (another *tsA* allele) occurs within *Hind* II + III fragment I (0.325 - 0.375 map position) (see Figure 6). Additional experiments using both techniques with other *tsA* mutants or A mutants resulting from deletions and insertions should better define the physical limits of the A cistron.

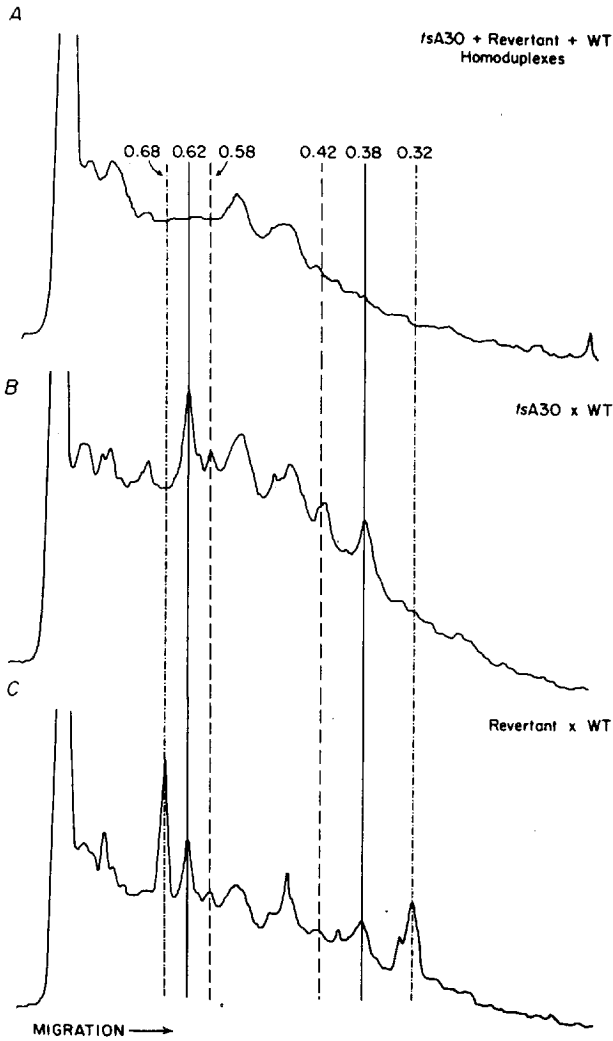


Figure 4. Cleavage by S1 nuclease of heteroduplexes prepared from *EcoRI* endonuclease-generated linear DNAs of *tsA30* with its wild-type parent and *tsA30-R* with the wild-type parent of *tsA30*. The S1 nuclease reactions contained 10 $\mu\text{g/ml}$ DNA. Samples of 0.75 μg DNA were applied to each gel. Electrophoresis was for 16 hours at 40 volts in agarose gels. The vertical lines drawn through the DNA bands serve to match pairs of fragments whose sizes add up to unit length SV40. (A) S1 nuclease-treated homoduplexes. (B) S1 nuclease-treated heteroduplexes formed from *tsA30* and wild-type DNAs. (C) S1 nuclease-treated heteroduplexes formed from *tsA30-R* and wild-type DNAs.

The S1 nuclease mapping procedure has been used to locate a second SV40 cistron. Heteroduplex molecules formed from either *Eco*RI or *Hpa*II endonuclease-cut *tsD202* and a spontaneous revertant, *tsD202-R*, were digested with S1 nuclease as indicated above (Figure 5). With homoduplexes of each DNA little or no specific cleavage was observed (Figures 5a and c). However, a fragment 0.91 SV40 fractional length was produced by S1 nuclease cleavage of the heteroduplex formed from *Eco*RI endonuclease-cut DNAs (Figure 5b) and a fragment 0.82 SV40 fractional length from the heteroduplex of *Hpa*II endonuclease-cut DNAs (Figure 5d). These results are consistent with the existence of a non-homology between *tsD202* and its revertant at 0.91 on the SV40 map.

In this instance also, the reversion appears not to have restored the wild-type sequence. This conclusion follows from the finding that there was a cleavage by S1 nuclease of the heteroduplexes formed from *tsD202-R* and wild-type DNAs (Figures 5b and d). Further, no specific S1 nuclease fragments were generated from the heteroduplexes formed between *Eco*RI endonuclease-generated linear DNAs from *tsD202* and its wild-type parent (Figures 5b and d). One must suppose, therefore, that the mismatch in the heteroduplexes prepared from *tsD202* and wild-type DNAs is not readily cleaved by S1 nuclease but that the heteroduplexes prepared from *tsD202-R* and *tsD202* DNAs and those from *tsD202-R* and wild-type DNAs contain a mismatch that can be cleaved by S1 nuclease.

Here, too, the data of Lai and Nathans (1975) substantiate our assignment of the D cistron to a region of the physical map containing the coordinate 0.91. They have found that *Hind* II + III fragment E (map position 0.86 - 0.945) contains the sequence modified by the *tsD202* mutation.

DISCUSSION

There seems little doubt that S1 nuclease can be used to map deletions and insertions (or, in fact, any change which produces single-strand segments in heteroduplex structures) in DNA. As with electron microscopic examination of heteroduplexes, the method permits accurate analysis of the location and size of these and other gross perturbations of DNA structure. Of particular importance, however, is the ability of the S1 nuclease mapping procedure to detect and locate deletions and insertions that are too small to be seen by electron microscopy. We are currently testing situations in which the deletions could be 25 base pairs or less.

A most promising but still provisional conclusion is that S1 nuclease can detect and map single base changes as well. We are presently cautious because, though reasonable, it is still unproven

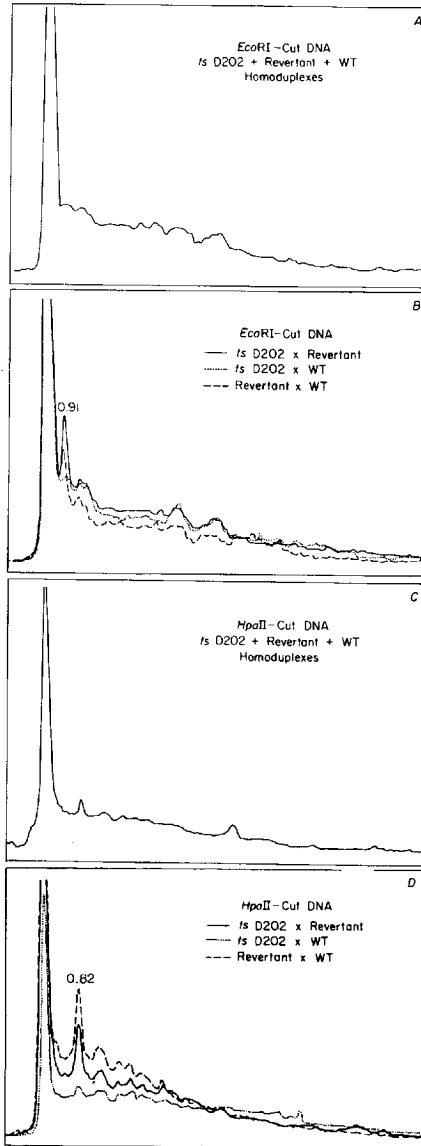


Figure 5. Cleavage by S1 nuclease of heteroduplexes: *tsD202* with *tsD202-R*, *tsD202* with its wild-type parent, and *tsD202-R* with the wild-type parent of *tsD202*. The S1 nuclease reactions contained 11.5 $\mu\text{g/ml}$ DNA. Samples of 0.75 μg DNA were applied to each gel. Electrophoresis was for 16 hours at 40 volts in agarose gels. (A) *EcoRI* endonuclease-generated, S1 nuclease-treated homoduplexes. (B) S1 nuclease-treated heteroduplexes prepared from *EcoRI* endonuclease-generated linear DNAs: —, *tsD202* with *tsD202-R*; ····, *tsD202* with wild type; - - - -, *tsD202-R* with wild type. (C) *HpaII*-generated, S1 nuclease-treated homoduplexes; (D) S1 nuclease-treated heteroduplexes from *HpaII* endonuclease-generated linear DNAs: symbols as in B.

that the temperature-sensitive DNAs differ from their revertant or wild-type DNAs by only single bases at the sites where S1 nuclease acts. Hopefully, experiments with phage DNAs having known base changes at specific locations can be used to resolve this uncertainty.

Our limited experience with the method has already indicated several shortcomings. Under the conditions used, cleavage of heteroduplexes with putative single base mismatches is slower and more limited in extent than that observed with deletion and insertion heteroduplexes; consequently, the signal (specific cutting at the mismatch) to noise (random cutting of the heteroduplex DNA) ratio is considerably lower in the former case. Several parameters are being studied to deal with this problem: the conditions required to ensure perfect reannealing in the formation of the heteroduplex substrates for S1 nuclease; the conditions which minimize the occurrence of regions of even transient single-strandedness (e.g., 'breathing' at regions of high AT content); the efficiency of S1 nuclease cleavage of mismatches involving different base pairs occurring within different sequences (e.g., A/G vs. T/C and the influence of neighbouring stretches of high or low GC content). A collateral problem related to the noise or non-specific cleavages is the possibility that sequence heterogeneity can result during propagation of a putative homogeneous DNA. With increasing experience we should also be better able to assign error limits to the lengths of the fragments and, therefore, to the map position of the mutations.

Even at its present stage of development, the S1 nuclease procedure has permitted us to map several changes in the A and D cistrons of SV40 (Figure 6). For example, at least two sites distinguish the base sequence of *tsA30* and its wild-type parent; one at 0.38 and the other at 0.42 (though possibly at 0.58). Only one difference was detected between the sequence in *tsA30* and its revertant *tsA30-R* and that occurs at 0.32 on the map. Consistent with this is the fact that *tsA30-R* differs from the wild-type at three map locations; 0.38, 0.42 and 0.32. Thus, *tsA30* is probably a double mutant (only one of the changes may be responsible for the *ts* phenotype) and the revertant mutation has occurred at a third site. However, all of the changes have occurred between the map coordinates 0.3 to 0.4, a region within which the *tsA30* and *tsA28* mutations have been located by marker rescue experiments (Lai and Nathans, 1975). Although we could infer the existence of a region of non-homology at map position 0.91 between *tsD202* and *tsD202-R* and between *tsD202-R* and wild-type DNAs, we were unable to detect a mismatch in the heteroduplexes formed between *tsD202* and wild-type DNAs. The reason for this is unknown and further experiments are needed. But here, too, the assignment of the *tsD202* mutation to map position about 0.91 agrees with Lai and Nathan's assignment of that mutation to *Hind* II + III fragment E.

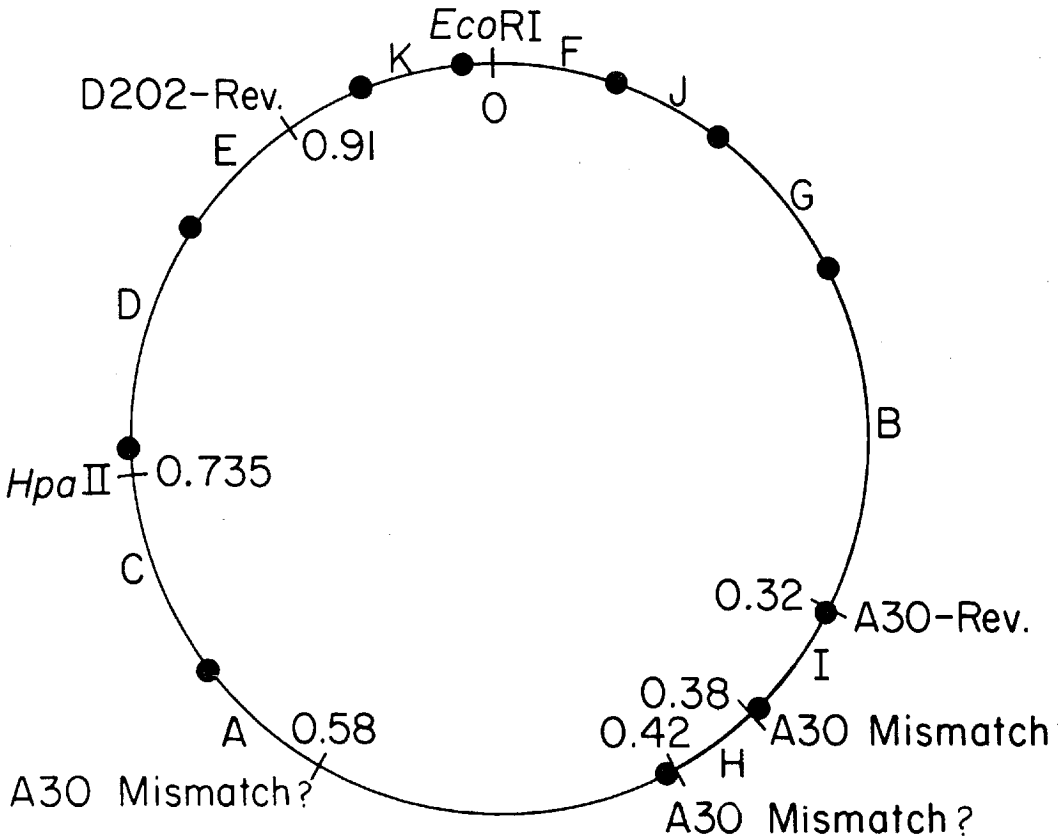


Figure 6. The locations of point mutations on the SV40 chromosome as determined by the S1 nuclease procedure. The solid circles on the map represent points at which *Hind* II + III endonucleases cut, and the letters designate the fragments produced by this cleavage (Danna et al., 1973).

Some mention should be made of several ramifications of the method. First, it provides an approach to genetic mapping in systems where genetic recombination is lacking or difficult to measure. Second, the procedure enables one to locate genotypic alterations which cause no detectable change in phenotype. Third, mutations of a variety of types can be physically mapped if the genome can be isolated in homogeneous form. We plan to investigate whether mutations or mismatches can be detected and mapped in other duplex nucleic acids, i.e., RNA and RNA-DNA hybrids. Finally, the ability of S1 nuclease to cleave a duplex DNA at a mismatch may make it possible to isolate discrete segments of a genome if that segment can be bounded by small deletions. For example, a heteroduplex between two different deletion mutants should yield three

fragments after S1 nuclease cleavage, one of which contains the segment between the two mutations.

ACKNOWLEDGMENTS

This work was supported in part by research grants from the U.S. Public Health Service (GM-13235-09) and the American Cancer Society (VC-23C). T.E.S. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. C.R. was supported by funds from the California Division of the American Cancer Society. P.W.J.R. is a Fellow of the Helen Hay Whitney Foundation.

REFERENCES

- Ando, T. 1966. A nuclease specific for heat-denatured DNA isolated from a product of *Aspergillus oryzae*. *Biochim. Biophys. Acta* 114: 158.
- Beard, P., J.F. Morrow and P. Berg. 1973. Cleavage of circular, superhelical SV40 DNA to a linear duplex by S1 nuclease. *J. Virol.* 12: 1303.
- Carbon, J. and P. Berg. 1974. A synthetic insert of poly(dA:dT) at the HpaII endonuclease cleavage site in SV40 DNA. *Proc. Nat. Acad. Sci. USA*. In press.
- Chou, J.Y. and R.G. Martin. 1974. Complementation analysis of SV40 mutants. *J. Virol.* 13: 1101.
- Danna, K.J., G.H. Sack and D. Nathans. 1973. A cleavage map of the SV40 genome. *J. Mol. Biol.* 78: 363.
- Germond, J.E., V.M. Vogt and B. Hirt. 1974. Characterization of the single-strand-specific nuclease S1 activity on double-stranded supercoiled polyoma DNA. *Eur. J. Biochem.* 43: 591.
- Greene, P.J., M.C. Betlach, H.M. Goodman and H.W. Boyer. 1974. The EcoRI restriction endonuclease. In *Methods in Molecular Biology* (W. Wickner, ed.). Marcel Dekker Inc. New York. In press.
- Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* 26: 365.
- Lai, C.J. and D. Nathans. 1974. Mapping temperature-sensitive mutants of SV40: Rescue of mutants by fragments of viral DNA. *Virology*. In press.
- , ———. 1975. Mapping *ts* mutants and deletion mutants of SV40. *Cold Spring Harbor Symp. Quant. Biol.* 39: In press.
- Mertz, J.E. and P. Berg. 1974a. Defective SV40 genomes: Isolation and growth of individual clones. *Virology*. In press.
- , ———. 1974b. Viable deletion mutants of SV40: Selective isolation using HpaII restriction endonuclease. *Proc. Nat. Acad. Sci. USA*. In press.

- Mertz, J.E., J. Carbon, M. Herzberg, R. Davis and P. Berg. 1975. Isolation and characterization of individual clones of SV40 mutants containing deletions, duplications and insertions in their DNA. *Cold Spring Harbor Symp. Quant. Biol.* 39. In press.
- Mertz, J.E. and R.W. Davis. 1972. Cleavage of DNA by R1 restriction endonuclease generates cohesive ends. *Proc. Nat. Acad. Sci. USA* 69: 3370.
- Peacock, A.C. and C.W. Dingman. 1968. Molecular weight estimation and separation of RNA by electrophoresis in agarose-acrylamide composite gels. *Biochemistry* 7: 668.
- Radloff, R., W. Bauer and J. Vinograd. 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. *Proc. Nat. Acad. Sci. USA* 57: 1514.
- Sharp, P.A., B. Sugden and J. Sambrook. 1973. Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* 12: 3055.
- Tegtmeyer, P. 1972. SV40 DNA synthesis: The viral replicon. *J. Virol.* 10: 591.
- Vogt, V.M. 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. *Eur. J. Biochem.* 33: 192.

The transfer and expression (transgenesis) of foreign genes in plant cells, reality and potential

Colin H. Doy

The use of the term transgenesis is discussed, including why it is preferred to transformation and transduction. Examples of transgenesis classified as DNA-mediated and phage-mediated transgenesis are presented and analysed with emphasis on present uncertainties and difficulties. If certain work with DNA-mediated transgenesis is accepted (but it remains a matter for controversy) then a transgenote of non-plant origin has already been bred into a seed bearing plant. In general, however, although the phenotype of DNA or phage-mediated transgenesis can persist indefinitely, there remains no unequivocal proof of inheritance, either nuclear or cytoplasmic. Possibilities for the future development of transgenesis are presented, including the creation of new transgenotes by using restriction and other enzymes. Caution is advised in applying transgenesis because of the possibility of creating transgenotes harmful to plants in general and therefore to the welfare of mankind.

The term transgenesis was coined in response to reaction to results obtained in conjunction with my then graduate student, Peter Gresshoff, and colleague Barry Rolfe. Phenotype is used to refer to a change in growth response or the appearance of a new molecule or function. We had changed the general phenotype of certain plant cells in defined culture from non-growth to growth by exposing the cells to bacterial viruses (specialized transducing phage) carrying specific bacterial (*Escherichia coli*) genes known to encode functions for converting the non-utilized carbon source to an available one. It seemed to us that the simplest (best?) explanation was that the bacterial genes were transferred into the plant cells and then transcribed and translated to a functional gene product using the machinery of the plant cells (Doy et al., 1972a,b). In a later example, where growth became non-growth due to the use of phage carrying a nonsense suppressor, only transcription was required (Doy et al., 1972b, 1973a,b).

Although we hoped that the new phenotypes might persist indefinitely (and they did) and that the transgenote might be inherited, we did not claim this. Our colleagues tended to either be extremely sceptical or else to claim for us that inheritance occurred and

moreover to ascribe their own image of mechanisms through the use of terms such as transformation and transduction. We were opposed to this and consequently the term transgenosis was introduced. We have defined transgenosis elsewhere (Doy et al., 1973a,b, 1974) but essentially it is meant to cover transfer and all processes, such as transcription and translation, needed for phenotypic expression, but does not automatically imply a claim for inheritance although, subsequently, inheritance may be proved. The term is also meant to emphasise that for many systems one is at present observing phenomena and offering hypotheses without any knowledge of mechanism except by inference against the background of general biological knowledge.

In the course of discussions in Australia and overseas, I have come to think of transgenosis as the transfer of genes by an asexual process, resulting in a change of phenotype but with the potential for a 'permanent' change of genotype to form a hybrid. The phenotype may appear before a process of inheritance is established, or may appear or disappear subsequent to genetical incorporation. I have therefore no objection if transgenosis is retained for examples where an inherited genotypic change is proven. In the long term, transgenosis may prove useful as a generalization for asexual transfer, expression and inheritance of genetical information between donor and recipient organisms (cells) widely separated by evolution.

The kind of reservations we apply to our work I think apply to many of the experiments concerned with the expression of prokaryote or viral genes in eukaryotic cells, especially of higher plants, and therefore are best described as transgenosis.

I should like to emphasise that the definition of transgenosis does not restrict one to a particular donor, vector, recipient or host. Thus the donor/vector could be nucleic acid, a virus, a cell or a protoplast and the recipient/host could be many kinds of protoplasts, cells, tissues or organisms, including those of animals as well as higher plants.

In this paper I shall not discuss protoplast fusion (but see Doy, 1974). Examples I shall discuss come under the general headings of DNA-mediated transgenosis and phage-mediated transgenosis.

DNA-MEDIATED TRANSGENOSIS

The work of Hess (1969a,b, 1970, 1973) is the one example of transgenosis I would have little reservation about describing as transformation in the bacterial geneticists usage (Hayes, 1964). Hess has mutants of *Petunia hybrida* and most of his experiments use DNA from red cyanidin strains (Hess, 1969a,b) to transgenose white

seedlings, white being a recessive character. Up to 25% of the treated plants gave red flowers. A control with homologous DNA gave about 5% of flowers that can be described as weakly red. Selfing of the most highly coloured transgenosed plants yielded seeds and subsequent progeny within which there was no segregation of the transgene. A similar result was obtained for differences of leaf shape controlled by a single gene mutation (Hess, 1970).

Several groups have claimed the persistence of bacterial DNA in higher plants but I shall concentrate on the work of Ledoux and his colleagues (Ledoux and Huart, 1961). Ledoux and his colleagues claim that treatment of seeds of various plants with bacterial DNA, often from *Micrococcus lysodeikticus*, leads to the covalent incorporation of donor DNA into the recipient plant's genome to form a double stranded end-to-end hybrid (Ledoux and Huart, 1961, 1968, 1972; Ledoux et al., 1971). This conclusion is based on the use of radioactive labelling techniques and the estimation of DNA buoyant densities before and after treatment, such as ultrasonication. Unlabelled *M. lysodeikticus* DNA ($\rho = 1.731$ g/cc) was used to treat successive generations of *Arabidopsis thaliana* grown with ^3H -thymidine. The labelled DNA was predominantly of a buoyant density intermediate between that of the donor and host ($\rho = 1.702$ g/cc). On shearing by ultrasonication, peaks of radioactivity were obtained with buoyant densities similar to the donor and host DNA. With each succeeding generation the intermediate buoyant density increased (F_0 DNA, $\rho = 1.702$; F_1 DNA, $\rho = 1.712$; F_2 DNA, $\rho = 1.719$; F_3 DNA, $\rho = 1.725$). This is a quite remarkable result since it says that more and more of the DNA synthesized in successive generations, helped by booster shots of bacterial DNA, is replicated from DNA of donor type. Ultrasonication of each sample, F_1 through F_3 , also gave DNA of buoyant densities similar to host and donor DNA.

I have seen similar data obtained by an Australian visitor to Ledoux's laboratory, Dr J.A.M. Brown (Brown, personal communication). As I understood it, in Brown's case the F_3 generation was followed by an F_4 without DNA treatment but continued exposure to ^3H -thymidine. Not only could radioactive donor and host DNA be recognized after sonication, but the radioactive distribution was mimicked by absorption at 260 nm. A control grown without DNA treatment contained only the normal plant DNA.

Workers in other laboratories seem unable to reproduce the results of Ledoux's group. Being negative, most of these studies have not been published. Hotta and Stern (1972) have published the results of a study where they attempted to repeat the early work of Ledoux but reached different conclusions. Hotta and Stern could not find donor DNA in the host nucleus but found peaks of inter-

mediate buoyant density only after X-irradiation of the host tissue or in response to environmental conditions which slowed cellular multiplication. They concluded that the formation of hybrid DNA was a pathological consequence of physiological mistreatment. If this is so, then there is an analogy with the 'stress DNA' (Quetier et al., 1961) which helps to confound interpretation of the crown gall phenomenon. In passing I should point out that the crown gall system strongly resembles phage-mediated transgenesis in that some workers believe that a phage transmits genetical information from the bacterium *Agrobacterium tumefaciens* into host plant cells. Although Hotta and Stern seem to refute Ledoux and his colleagues, they did, in some conditions, mimic their results, in that they found peaks of intermediate density and a base composition intermediate between donor and host DNA and could, by ultrasonication, obtain peaks of donor and host DNA.

If one accepts Ledoux's conclusions, then clearly it is relatively easy to introduce large amounts of double stranded bacterial DNA into the plant nuclear genome. Moreover this replicates and becomes a permanent part of the genome and is inherited through the seeds. This latter is the most desirable target if transgenesis is to be used to breed novel plants.

Ledoux and his colleagues have also been able to restore prototrophy to nutritional mutants of *Arabidopsis thaliana* by treating them with DNA of various kinds (Ledoux et al., 1972). The mutants concerned are thiamine auxotrophs, a class of mutants well established in the literature, and a tryptophan-requiring mutant (although probably not deficient in tryptophan biosynthesis) which is not available to other workers. The tryptophan mutant is leaky and a subject of controversy, and I shall not comment further. The thiamine mutants can be restored to prototrophy using DNA of animal and bacterial origins. Correction is a fairly rare event, but prototrophy is inherited and there is no segregation. There is no evidence that the plant nucleus is the site of the effective transgenote.

Once again, these results are surrounded with controversy. A control that would be informative is one using DNA carrying a marker closely linked with the thiamine gene of the transgenote. Subsequent detection of this second but unselected marker in plants prototrophic for thiamine would be very strong evidence in favour of transgenesis and transformation in the sense of the bacterial geneticist.

The ease with which the Ledoux technique introduces foreign DNA into the plant genome is also a disadvantage since much of the genetical information is unwanted and may even be harmful. In fact plants of abnormal appearance are produced. Abnormality may be a

consequence of transgenesis of unwanted genes or be indicative of mutagenic effects of the foreign DNA. Mutagenesis could also account for the recovery of prototrophs. Recovery of a non-selected transgenote would help to refute this possibility.

PHAGE-MEDIATED TRANSGENOSIS

I shall summarize our published work (Doy et al., 1972a,b, 1973a,b,c), briefly discuss the work of others and indicate the trend of more recent studies. We looked for a situation of non-growth of haploid plant cell lines kept in callus culture. Situations of non-growth for haploid *Lycopersicon esculentum* (tomato) callus were media containing either galactose or lactose as the sole general carbon source. It was found that specialized transducing phage carrying the appropriate bacterial genes for enzymes converting these sugars into more available carbon sources, for example glucose, were successful in initiating general growth. The phage in question were λ pgal and ϕ 80plac and the technique was simply to drop purified phage onto a small piece of callus and wait. The callus had to be on non-differentiation medium rather than differentiation medium which suggests that the physical and/or metabolic state of the recipient cells is important.

One of the negative controls (no growth of treated callus) or galactose was a phage carrying a nonsense mutation in gal (λ pgal⁻, amber UAG). This negative result suggested that the plant cells could not efficiently correct UAG. It was reasoned that UAG codons might be vital to the plant cell, for example, if they participated as full stops during translation. Suppression of nonsense would therefore be harmful. Accordingly we took both haploid tomato and *Arabidopsis thaliana* callus growing on optimal glucose medium and added ϕ 80supF⁺. The *E. coli* mutant gene, supF⁺, specifies a t-RNA able to insert tyrosine at UAG. In the event, ϕ 80supF⁺ killed the calluses even at the relatively low initial phage titre of 100 plaque forming units per plant cell.

In all the above experiments, neutral phage, that is phage without genes specifying relevant products, for example ϕ 80 added to callus on glucose medium, appeared harmless. However, I found that after three rounds of sub-culture, using many samples from each callus, the growth rate of many inocula declined to about half that of controls. Continual sub-culture on glucose medium of callus exposed once to ϕ 80 generally leads to a rapid decline in growth rate and eventual death. This I think is not unexpected since there seems no reason why transgenesis should be confined to the 'wanted' *E. coli* genes carried within the phage genome. Indeed there is evidence for transgenesis of phage T3 genes (Carlson, 1973 and see below).

The delay in affecting growth can be rationalized as follows. On optimal growth medium the plant cells initially have a selective advantage over transgenosed phage genes. However, as callus growth declines there is a chance that replication and expression of phage genes could continue as a parasitic 'disease'. In these examples of accelerated death there is no evidence for the presence of culturable micro-organisms. When examined for phage the original titre of plaque forming units falls to zero over the first 40 or more days (assays of callus and, separately, the medium).

As part of the follow-up to growth on lactose due to *lac* gene transgenesis, we looked for evidence of β -galactosidase synthesis. Under our assay conditions the plant cells normally have only a very low level of β -galactosidase (or an analogous activity) but, by contrast, some calluses of those presumed by growth to be transgenosed have very high levels of β -galactosidase and this activity is protected by antiserum to the known *E. coli* enzyme specified by *lac* gene Z.

No work has continued on *gal* since the transgenosed material grows so slowly. Calluses transgenosed for *lac* are much better and some are still growing on lactose after more than two years.

All calluses that showed high levels of gene Z β -galactosidase (probably synthesized during a burst of expression, see Doy et al., 1973c) eventually declined and died at various rates despite comprehensive efforts to revive them on media variations, including those that normally favour differentiation.

Those calluses that survive transgenesis for gene Z generally contain only low levels of β -galactosidase and this activity could not be shown to be unambiguously positive in the immunological test for the *E. coli* enzyme. Thus these calluses are similar to the occasional control that grows slowly on lactose without specific phage treatment. However, the presumptive transgenosed calluses differed from these other calluses in that they could be established and greened on lactose-supplemented differentiation medium in the light. Thus exposure to the relevant phage seemed a prerequisite for survival on transfer. Most subsequent work has been done on these calluses which grow well and indefinitely on lactose.

No experiments have been attempted with nucleic acid hybridization for despite much attention, hybridization studies have not provided unequivocal answers concerning the role of bacterial and phage genes in crown gall. Further, the hybridization experiments of Merrill et al. (1971) did not lead to general acceptance of the expression of *gal* in human cells.

Accordingly I attempted to generate whole homozygous diploid plants for classical genetical and biochemical studies from transgenosed calluses. All attempts have failed to date. It is possible to get the plant cells to a stage of making chloroplasts and mitochondria, as shown by electron microscopy, but not to go further. This is not necessarily a consequence of transgenesis since the original haploid tomato cell lines that once gave plantlets no longer do so. Anyone familiar with plant cell culture will recognize that this is not a unique problem and represents an area which requires much basic research.

Subcultures of the haploid tomato cell line used in this laboratory have now yielded suspension cultures which grow well on glucose and sucrose but not on lactose (up to 10%). This confirms the conclusions with callus cultures and the suspension cultures may provide better material for further experiments if the conditions are conducive to transgenesis.

Two other laboratories have published results that have many points in common with ours. Carlson (1973) found that two gene products of phage T3 were expressed in protoplasts of *Hordeum vulgare* (barley). These are S-adenosyl methionine cleaving enzyme and RNA polymerase encoded in the early region of the T3 genome. A nonsense mutation in the T3 RNA polymerase gene prevented expression of the enzyme. This is the same as our finding with *gal*.

Johnson et al. (1973) found that suspensions of *Acer pseudoplatanus* (sycamore) grew slowly on lactose when exposed initially to λ plac5. In the absence of exposure to phage, or on exposure to λ , no growth occurred. These authors did not find increased levels of β -galactosidase or successfully demonstrate the presence of the gene Z product by immunology. Their results reflect our findings in the first few weeks of transgenesis of *lac* and for the eventual long term survivors on lactose. A cell suspension system should be ideal for selecting the best survivors and these, whether callus or cell suspensions, may depend on a low level balanced expression of transgenesis. The question of whether long term survivors, containing low levels of β -galactosidase, are expressing the transgenote has not been answered either by me or by Smith's group (Johnson et al., 1973). In my hands judgment is confounded in that a few controls always did grow on lactose which means that this phenotype could arise spontaneously by epigenetic or mutational change at any time in cultures for which growth was initiated by transgenesis. If this led to a more efficient supply of a usable carbon source, or if transgenosed cells always die with the present phage vectors, then these phenotypes and genotypes would be selected and growth would, from then on, be nothing to do with transgenesis. Another alternative to direct

expression is that *lac* genes in a transgenote somehow turn on new plant functions. This cannot be a non-specific response to phage since *lac* genes must be present.

Stress has already been mentioned as a possible factor in DNA-mediated transgenesis. Similarly phage-mediated transgenesis is always started from a situation of stress, for example, cultures that do not normally grow, or from an inoculum of callus just transferred.

In summary, phage-mediated transgenesis has been achieved in three laboratories and in this it differs from DNA-mediated transgenesis. Two of these laboratories deduce that a phage gene product is synthesized in plant cells or protoplasts. The two laboratories working with systems that result in growth find that growth and cell-division persists indefinitely. All workers agree that success requires that the vector carry a relevant and functional gene.

THE POSSIBLE FUTURE OF TRANSGENOSIS

Transgenesis has two aims, basic research and applied objectives. For some basic research we are not concerned whether the foreign genes persist indefinitely or are inherited. Useful information can be gained solely if expression occurs for a time. This was in fact our initial objective. For some basic and all applied objectives the transgenote must be inherited, preferably within the nuclear genome or by integration with an existing mechanism of cytoplasmic inheritance. Part of the mechanics of achieving inherited transgenesis involves selection and this may require the use of forcing markers in the transgenote, e.g., mutations for drug resistance, preferably linked closely to the desired transgenote. Some factors affecting applied objectives have been discussed elsewhere (Doy, 1974).

For long term objectives I am inclined to reject DNA-mediated transgenesis as practised by Ledoux and also phage-mediated transgenesis in its present form. Phage might be modified to minimise the transfer of unwanted genes, for example, by mutations or deletions of the positive control N gene of λ , or by the substitution of wanted for unwanted genes. I understand that more than 90% of the λ genome could be replaced.

The concept of gene substitution leads to what I think is the real future of transgenesis and that is the application of restriction and other enzymes to make new kinds of hybrid DNA, including viruses and plasmids. I shall not detail this area, since Hogness and Berg have described elsewhere in this symposium work with *Drosophila* and SV40 virus which elegantly illustrates the potential,

as do recent papers by Nathans (1974) and Boyer (1974). Using these techniques new combinations of genes are created, including combinations from diverse biological origins.

These new DNAs could serve as transgenotes for either DNA-mediated or viral-mediated transgenesis. The inclusion of homologous plant DNA in the potential transgenote would maximise the possibility for the donor genes to become integrated into plant chromosomes. The inclusion of DNA from a phage such as λ gives the possibility of building copies of the transgenote in *E. coli*.

A further intriguing possibility arises from the similarity of plant and bacterial metabolism. One could join random pieces of plant DNA to phage DNA (e.g. λ) and then, by using an *E. coli* deleted for a particular gene, look specifically for the λ -plant DNA hybrid carrying the analogous plant function. This assumes reciprocity of transgenesis (eukaryote genes in a prokaryote). The desired hybrid would be selected and multiplied up in *E. coli* to give a high titre of the transgenote for putting back into the original plant species or another. This approach might also be used for mapping the selected plant gene in the original genome.

One therefore has the potential for selectively changing plant genes or increasing the number of copies of desirable plant genes. One might also seek to mutate plant genes and carry out the selection in *E. coli* before applying them back to plants. Clearly the possibilities are enormous. The technology from the side of the microbiologist and molecular biologist is available, the limiting factors are the plant systems. The general handling of plant cells in culture, cloning, the ability to direct development and differentiation and to select mutants, is still poorly understood, although material from some plants can be manipulated much at will.

I shall conclude with a word of caution. The result with $\phi 80_{supF}^+$ worried me immensely since we killed the plant cells. Luckily only cells on non-differentiation medium were susceptible. The use of restriction enzymes and transgenesis leads to possibilities of either knowingly or unknowingly creating dangerous forms of DNA. I have said nothing about plant viruses, since at present I think they add nothing, but the manipulation of their genomes (most are, however, RNA viruses) might be exceedingly dangerous.

REFERENCES

- Boyer, H. 1974. Site-specific endonucleases in the analysis and restructuring of DNA molecules. Proc. workshop on genetic improvement of seed proteins. *Nat. Acad. Sci. USA*. In press.
- Carlson, P.S. 1973. The use of protoplasts for genetic research. *Proc. Nat. Acad. Sci. USA* 70: 598.

- Doy, C.H. 1974. Asexual approaches, including transgenosis and somatic cell hybridization, to the modification of plant genotypes and phenotypes. Proc. workshop on genetical improvement of seed proteins. *Nat. Acad. Sci. USA*. In press.
- Doy, C.H., P.M. Gresshoff and B.G. Rolfe. 1972a. Transgenosis of the galactose operon from *Escherichia coli* to cultured cells of haploid *Lycopersicon esculentum* (tomato). *Proc. Aust. Biochem. Soc.* 5: 3.
- _____, _____, _____. 1972b. Transfer and expression (transgenosis) of bacterial genes in plant cells. *Search* 3: 447.
- _____, _____, _____. 1973a. Biological and molecular evidence for the transgenosis of genes from bacteria to plant cells. *Proc. Nat. Acad. Sci. USA* 70: 723.
- _____, _____, _____. 1973b. Transgenosis of bacterial genes from *Escherichia coli* to cultures of haploid *Lycopersicon esculentum* and haploid *Arabidopsis thaliana* plant cells. In *The Biochemistry of Gene Expression in Higher Organisms* (J.K. Pollak and J. Wilson Lee, ed.) p.21. Australia and New Zealand Book Co., Sydney.
- _____, _____, _____. 1973c. Time-course of phenotypic expression of *Escherichia coli* gene Z following transgenosis in haploid *Lycopersicon esculentum* cells. *Nature New Biol.* 244: 90.
- Hayes, W.H. 1964. *The Genetics of Bacteria and their Viruses*, pp. 7, 55, 523 and 524. Blackwell, Oxford.
- Hess, D. 1969a. Versuche zur Transformation au hoheren Pflanzen: Induktion und konstante Weitergabe der Anthocyansynthese bei *Petunia hybrida*. *Z. Pflanzenphysiol.* 60: 348.
- _____. 1969b. Versuche zur Transformation au hoheren Pflanzen: Wiederholung der Anthocyaninduction bei *Petunia* und erste Charakterisierung der transformierenden Prinzips. *Z. Pflanzenphysiol.* 61: 286.
- _____. 1970. Versuche zur Transformation au hoheren Pflanzen: Mogliche Transplantation eines Gens fur Blattform bei *Petunia hybrida*. *Z. Pflanzenphysiol.* 63: 461.
- _____. 1973. Transformationversuche au hoheren Pflanzen: Untersuchungen zur Realisation der Exosome-Modells der Transformation bei *Petunia hybrida*. *Z. Pflanzenphysiol.* 68: 432.
- Hotta, Y. and H. Stern. 1972. Uptake and distribution of heterologous DNA in living cells. In *Informative Molecules in Biological Systems* (L. Ledoux, ed.) p.176. North Holland, Amsterdam, London.
- Johnson, C.B., D. Grierson and H. Smith. 1973. The expression of λ plac5 DNA within cultured cells of a higher plant. *Nature New Biol.* 244: 105.

- Ledoux, L. and R. Huart. 1961. Sur la possibilite d'un transfert d'acides ribo- et desoxyribonucleiques et de proteins dans les embryons d'orge en croissance. *Arch. Intern. Physiol. Biochim.* 69: 598.
- _____, _____. 1968. Integration and replication of DNA of *M. lysodeikticus* in DNA of germinating barley. *Nature* 218: 1256.
- _____, _____. 1972. Fate of exogeneous DNA in plants. In *Uptake of Informative Molecules by Living Cells* (L. Ledoux, ed.) p. 254. North Holland, Amsterdam, London.
- Ledoux, L., R. Huart and M. Jacobs. 1971. Fate of exogenous DNA in *Arabidopsis thaliana* I. Translocation and integration. *Eur. J. Biochem.* 23: 96.
- _____, _____. 1972. Fate of exogenous DNA in *Arabidopsis thaliana* II. Evidence for replication and preliminary results at the biological level. In *Informative Molecules in Biological Systems* (L. Ledoux, ed.) p.159. North Holland, Amsterdam, London.
- Merril, C.R., M. Geier and J.C. Petricciani. 1971. Bacterial virus in gene expression in human cells. *Nature* 233: 398.
- Nathans, D. 1974. Site-specific endonucleases in the analysis and restructuring of DNA molecules. Proc. workshop on genetic improvement of seed proteins. *Nat. Acad. Sci. USA*. In press.
- Quetier, F., T. Huguet and E. Guille. 1961. Induction of crown gall: Partial homology between tumour cell DNA, bacterial DNA and the GC rich DNA of stressed normal cells. *Biochem. Biophys. Res. Comm.* 34: 128.

Characteristics of bacterial gene expression in human fibroblasts

Carl R. Merrill, Mark R. Geier and Barry G. Rolfe

Cultured human cells from patients with a congenital lack of the enzyme α -D-galactose-1-phosphate uridyl (GPU) transferase, which is necessary for the metabolism of the sugar galactose, can have their enzyme deficiency partially restored by infection with a λ transducing phage carrying bacterial genes for galactose metabolism. Studies of the galactosemic human cells which have undergone successful transgenesis for the bacterial transferase activity show that this enzyme is capable of functioning to some extent in the galactose metabolic pathway of human cells and that transgenosed cells are now capable of surviving between two to three weeks longer than control cells in a medium containing galactose as sole hexose source.

Human fibroblasts from a patient with galactosemia were used as a model system to show that bacterial genes carried by a virus are capable of functioning in a human cell milieu and as an experimental approach to gene therapy using genes carried by viruses (Merril et al., 1971; Geier and Merrill, 1972; Rogers and Pfuderer, 1968). These experiments utilized a λ virus carrying the complete galactose operon $\lambda pgal$ ($K^+T^+E^+$) and a virus $\lambda pgal$ ($K^+T^-E^+$) with an amber point mutation in the transferase gene as a control. The galactosemic cells CCL-72 (from the American type culture collection) had no detectable D-galactose-1-phosphate uridyl transferase activity. However, after infection with $\lambda pgal$ ($K^+T^+E^+$), transferase activity ranged from 5.9% to 29% of that found in normal cells (Merril et al., 1971). Purified $\lambda pgal$ ($K^+T^+E^+$) DNA also gave transferase activity in the galactosemic cells. In these experiments cells infected with $\lambda pgal$ ($K^+T^-E^+$) DNA produced no transferase. The use of DNA-RNA hybridization techniques also revealed the presence of λ specific RNA following infection of human galactosemic cells with this virus (Geier and Merrill, 1972).

Since the publication of our initial observations, successful demonstrations of bacterial gene expression in several plant cell systems have been reported. $\lambda pgal$ was shown to rescue tomato calluses on galactose selection media and that other bacterial genes could also be successfully introduced by λ transducing phages into various experimental plant-cell systems (Doy et al., 1973b,c;

Johnson et al., 1973). Progress and problems in this field have recently been reviewed (Merril et al., 1972; Merrill, 1974).

TRANSFERASE ACTIVITY IN DIFFERENT GALACTOSEMIC CELL LINES

To examine whether or not our original results were unique to the CCL-72 cell line, we have extended our studies to three additional galactosemic cell lines GM52, GM53 and GM54, which have no detectable transferase activity. Transferase activity (Table 1)

TABLE 1

Restored transferase activity in human galactosemic cell lines

	Normal human cells	Galactosemic cell lines			
		CCL-72	GM-52	GM-53	GM-54
Transferase activity	+	-	-	-	-
Range of restored transferase activity in λ pgal ⁺ infected cultures (percent of normal cell level)		4.3-21%	0.6-35%	0.3-10%	0.1-1.2%
Overall success of λ pgal ⁺ experiments to restore transferase activity in human galactosemic cells		68%			

A survey of 28 independent experiments in which galactosemic fibroblasts were infected with λ pgal (K⁺T⁺E⁺). The virus was grown vegetatively, precipitated with polyethylene glycol and banded in CsCl as described previously (Merril et al., 1971). Line CCL-72 was obtained from the American Type Culture Collection, Rockville, Md, U.S.A., lines GM-52, GM-53 and GM-54 were obtained from the Institute for Medical Research, Genetic Mutant Cell Culture Repository, Camden, New Jersey, U.S.A. Cell culturing and phage infection was carried out as described elsewhere (Merril et al., 1971). The enzyme assays were carried out as described by Merrill et al. (1971) except that the glycine buffer used was 0.01 M and the alcohol was removed from the ¹⁴C-galactose-1-P by bubbling nitrogen through it prior to its use. The normal human fibroblasts (Led-130) produce about 35 n moles UDP-galactose under the enzyme assay procedures used. In all experiments the percentage of restored transferase activity is normalized against the Led-130 cell line values.

could be demonstrated only in each of these cell lines following infection with λ_{pgal} ($K^+T^+E^+$) virus. The range of the restored transferase activity varies within each cell line and might therefore reflect the type of galactose transferase mutation carried by each cell line. Out of a total of 28 independent experiments carried out with these three additional cell lines, measurable transferase activity was obtained in 68% of the experiments only in those cells previously infected with the λ_{pgal}^+ virus. No activity was obtained in uninfected cells or cells infected with λ_{pgal}^- virus. This latter finding means that the four human galactosemic cell lines examined have been unable to suppress the amber *UAG* nonsense mutation carried in the T (transferase) gene of λ_{pgal}^- phage. The cell cultures used in these experiments were constantly monitored to rule out contamination by bacteria and mycoplasma. As an additional control, enough bacterial transferase enzyme was added to a flask of cells to give 128,000 cpm of UDP-Gal formed in 25 min in our enzyme assay system. No counts above the normal background of 45 cpm were observed, demonstrating that contaminating enzyme could not be passively carried into the cells. Similarly, if the bacterial enzyme plus wild-type virus λ (non-transducing) is added to a flask of cells, no enzyme is detected. These results indicate that our original findings were not unique to the CCL-72 cell line. The wide variation in amount of enzyme observed following infection with λ_{pgal} ($K^+T^+E^+$) is not yet understood, nor are the failures (32% of the experiments) to obtain enzyme following infection with the T^+ virus. However, this variation is reminiscent of the early work on transformation in *Pneumococcus* and the phenomenon of 'cell competence' (Hotchkiss, 1954). Our initial experiments to test this possibility indicate that the most reproducible results were obtained by infecting rapidly growing (mid-logarithmic phase) cultures with high titres of transducing virus. Conversely, cell cultures approaching stationary phase of growth or newly sub-cultured cells did not give consistent transferase activity. Moreover, this latter finding probably explains the inability of Thilly et al. (1973) to transgenose thymidine kinase-deficient Chinese hamster cells with a phage carrying the thymidine kinase gene of *Escherichia coli* as they used newly sub-cultured cells left overnight before phage infection.

PROPERTIES OF THE BACTERIAL TRANSFERASE ENZYME IN HUMAN CELLS

Preliminary transferase inhibition studies were carried out on crude preparations of bacterial, normal human mammalian, and enzyme formed after λ_{pgal} ($K^+T^+E^+$) infection of human galactosemic fibroblasts (transgenosed enzyme, T.E.) (Table 2). Mixing experiments were also done to show that the results were due to the properties of the respective enzymes rather than to the cellular milieu in which the enzymes were assayed. The results indicated that of the four

inhibitors studied, the T.E. enzyme appeared 'bacterial like' in the presence of galactose; however, with respect to heat stability (45°C for 20 min), it is 'mammalian like'. With the inhibitors 2-deoxy-glucose-6-P and xylose-1-P, the T.E. enzyme gave an intermediate response compared with the normal human and the bacterial enzyme. There are at least three models which could explain these results. First, although a bacterial enzyme is made in the human cells, some errors in transcription and/or translation could give it intermediate properties. A second possibility is that the enzyme formed is a hybrid mammalian-*E. coli* enzyme. Some galactosemic cell lines have been shown to be CRM positive, that is, they have a defective transferase which is antigenically similar to normal transferase (Tedesco and Mellman, 1971). The subunits of such an enzyme might interact with the bacterial protein synthesized following the infection with a λ pgal⁺ virus. In fact, interallelic complementation studies on hybrid human galactosemic cells suggest that functional

TABLE 2
Differential enzyme inhibition study

Inhibitors	Percent activity remaining after treatment			
	Normal human cell lysate	<i>E. coli</i> cell lysate	<i>E. coli</i> enzyme in GM-52 cell lysate	λ pgal (K ⁺ T ⁺ E ⁺) infected GM-52 cell lysate
2-deoxy-glucose-6-P	9	82	64	35
Xylose-1-P	23	71	71	31
Galactose	17	88	90	96
Heat 45°C for 20'	<1	133	116	<1

The inhibitors were used at 30 mM final concentration in the enzyme assays. The data are shown as percent of uninhibited activity. Percent of uninhibited activity (cpm with inhibitor/cpm without inhibitor) x (100). The assays were carried out as described by Merrill et al. (1972). The non-galactosemic human cells were cell line MM and were a gift from Dr B.W. Uhlenhof, while the *E. coli* strain N205, from the NIH collection, was a GPU-transferase positive strain. GM-52 is a galactosemic human cell line from the Institute for Medical Research, Genetic Mutant Cell Repository, Camden, N.J., U.S.A. Each determination was the result of two kinetic studies and the value given is the average of these determinations.

hybrid molecules can be formed (Nadler et al., 1970). Finally, the third possibility is that the physical milieu in the mammalian cell is such that the tertiary structure altered during the enzyme synthesis. The data are at this time insufficient to narrow the model to a unique solution; however, the intermediate behaviour of this enzyme in cells infected with the transducing phage is further evidence against the enzyme being merely the result of bacterial enzyme contaminant, or induction of the normal mammalian enzyme.

METABOLIC EFFECT OF *E. COLI* OPERON IN GALACTOSEMIC CELLS

To determine whether the introduced bacterial genes can function effectively *in vivo* within the mammalian cell we have studied the liberation of $^{14}\text{CO}_2$ from galactose-1- ^{14}C used as sole hexose source in the growth medium. The conversion of galactose-1- ^{14}C to $^{14}\text{CO}_2$ was first used to measure the ability of galactosemic and normal fibroblasts to utilize galactose. Normal human fibroblasts were found to convert about twice as much galactose to CO_2 as do galactosemic fibroblasts (Figure 1). The utilization of galactose by galactosemic cell strains appears to be characteristic of galactosemic cells and is consistent with observations that some galactosemic individuals are able to convert the carbon-1 of intravenously administered galactose into carbon dioxide (Segal et al., 1962). In fact, liver cells from one galactosemic patient were found to catabolize gal-1- ^{14}C with an efficiency of 75% of the normal controls (Topper et al., 1962).

Actually four alternative pathways for galactose utilization by galactosemic cells have been proposed. They involve the conversion of: (1) galactose-1-phosphate (gal-1-P) to uridine diphosphate galactose (UDP-gal) by a pyrophosphorylase (Isselbacher, 1957); (2) galactose to D-xylulose (Cuatrecasas and Segal, 1966); (3) galactose to 6-phosphogalactonate (6-P-gal) (Inouye et al., 1962); and (4) galactose to galactonolactone and galactitol (Kalckar et al., 1959; Segal et al., 1965). It has also been suggested that some cultured galactosemic cells have residual transferase activity (Russell, 1968); however, we have been unable to detect any transferase activity in any of the galactosemic cells which we have tested (CCL-72, GM-52, GM-53 or GM-54).

To test the effect of λ transducing viruses on human galactosemic fibroblasts, a comparison was made of replicate flasks of strain CCL-72 cells inoculated with either λ^+ (non-transducing) or λpgal^+ and non-infected cultures. After gal-1- ^{14}C was added, the flasks were sealed and periodically monitored for $^{14}\text{CO}_2$ production (Figure 2). In every case where we have even exposed galactosemic cells to λpgal^+ infection we have been able to measure more $^{14}\text{CO}_2$ produced than in non-infected cells, cells infected with λ^+ or with

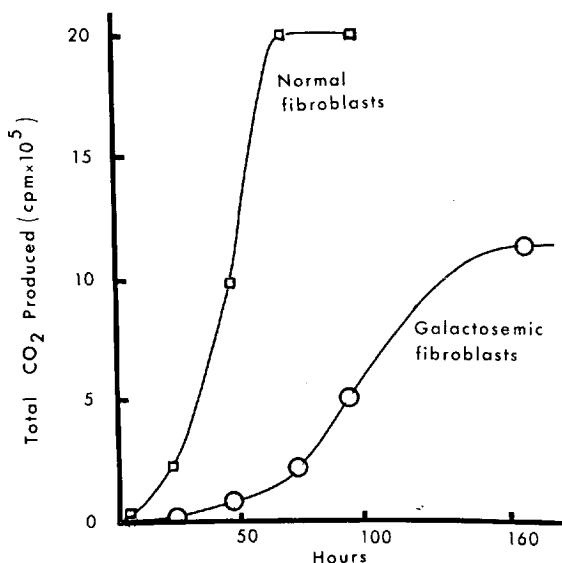


Figure 1. The conversion of galactose-1- ^{14}C to $^{14}\text{CO}_2$ by normal and galactosemic fibroblasts. CO_2 production was monitored as described elsewhere (Petricciani et al., 1972). The cells were washed twice with phosphate buffered saline. The hexose free MAB 87/3 containing 10% dialysed foetal calf serum and ^{14}C -labelled galactose (1.05×10^7 cpm, 98 mM) was added to each flask and silicone rubber stoppers were inserted. The flasks were then incubated at 37°C and $^{14}\text{CO}_2$ production was monitored by removing measured gas samples with an oiled syringe. The $^{14}\text{CO}_2$ was then trapped in phenethylamine columns absorbed to chromsorb. The columns were then tapped into scintillation counting vials and counted in fluid containing TLA and toluene (Beckman). The gal-1- ^{14}C used throughout these experiments was obtained from New England Nuclear (Lot #379-0643 0.938 mg/2.5 ml; 3.77 mg/mCi; 48.1 mCi/mM). There was no measurable radioactive volatile product from this lot of galactose which could be trapped in the phenethylamine column even following incubation for six days at 37°C in the medium used in these experiments. Tests with hexokinase revealed less than 1% contamination with other hexoses; this was corroborated by thin layer chromatography on cellulose with an ethylacetate:pyridine: H_2O (2:1:2) system. Chromatography revealed the presence of no hexose other than galactose, and no significant radioactivity in other areas of the chromatogram. Cells used here: human galactosemic fibroblast strain CCL-72 (passage 16) was obtained from the American Type Culture Collection, Rockville, Maryland. Strain Led-130, a human fibroblast derived from foetal lung, was kindly provided in passage 14 by Miss Roslyn Wallace of Lederle Laboratories, Pearl River, New York. Cells were cultured as monolayers with medium MAB 87/3 containing penicillin (100 u/ml), regular insulin (0.08 u/ml) and 10% foetal calf serum. Dialysed foetal calf serum was substituted during the $^{14}\text{CO}_2$ production experiments. Tests for bacterial contamination of the cells included inoculation of spent fluids into trypticase soy-broth and thioglycollate media and holding for 10 days at room temperature and 37°C . The cell strains were also periodically monitored for Mycoplasma contamination. All such tests were negative in the experiments reported here.

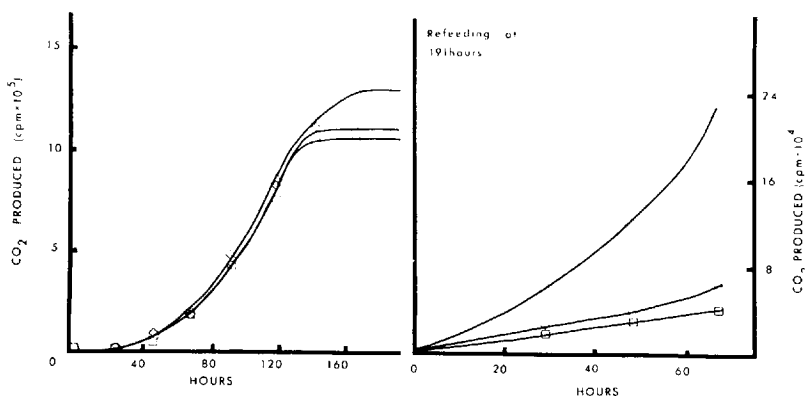


Figure 2. The effect of λ gal (K⁺T⁺E⁺) on galactose metabolism in the CCL-72 galactosemic fibroblast line. ○ λ gal⁺, ◇ λ, □ uninfected cells.

λ gal⁻ phage. This increased ¹⁴CO₂ production has ranged from 150,000 to 300,000 cpm above the control cells. These findings have been consistently observed in numerous independent experiments. Furthermore, the greater metabolic potential of the cells exposed to λ gal⁺ compared with uninfected cells or those exposed to wild-type λ is best illustrated by re-feeding each group of cells at the point at which CO₂ production has reached a plateau (Figure 2, Exp. A). Only in those cell cultures that have received λ transducing viruses is a significant level of liberated ¹⁴CO₂ detected, which can amount to nearly 20% of that detected before re-feeding. The control cultures, by comparison, produce only about 2% ¹⁴CO₂ after re-feeding (Figure 2, Exp. A re-feeding at 191 hours). The amplification of the differences in ¹⁴CO₂ production probably reflects a combined higher ¹⁴C survival rate and increased ability to convert gal-1¹⁴C to ¹⁴CO₂ by the cells exposed to λ gal⁺.

Metabolic enhancement of galactose utilization has also been observed in other galactosemic cell lines (Figure 3).

We have tried to use this fact to see if we could monitor the cells more effectively and determine the optimum time when to assay the cell cultures for the bacterial enzyme activity. Figure 4 shows that we have been able to detect transferase activity in cell cultures during the time when they were exhibiting maximum enhanced galactose utilization (Figure 4). Enzyme assays carried out on cell cultures at times before the point of maximum CO₂ production usually failed to show any activity. It seems, therefore, to obtain the optimum time when to assay cell cultures for bacterial transferase

activity, measurements should be carried out on the $^{14}\text{CO}_2$ liberated from parallel virus-infected cultures.

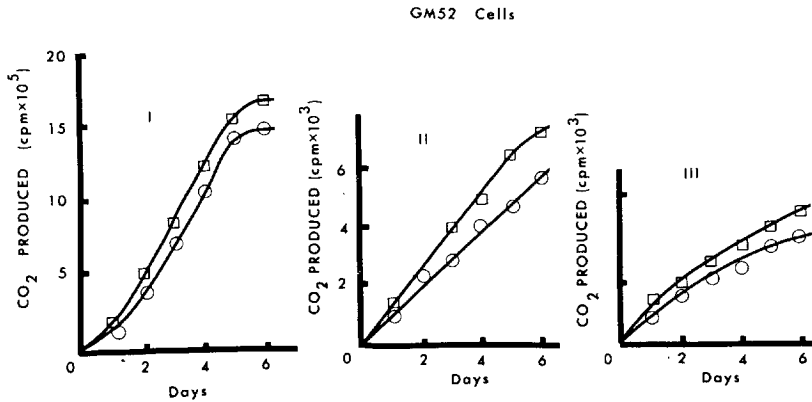


Figure 3. The effect of λpgal ($\text{K}^+\text{T}^+\text{E}^+$) on galactose metabolism in the GM-52 galactosemic fibroblast line. I, 0.05 μ moles galactose-1- ^{14}C . II, 0.05 μ moles galactose-1- ^{14}C + 560 μ moles cold galactose. III, 0.05 μ moles galactose-1- ^{14}C + 1100 μ moles cold galactose. \square λpgal^+ ; \circ uninfected cells.

Thus λpgal ($\text{K}^+\text{T}^+\text{E}^+$) can increase the ability of galactosemic cells to utilize galactose. This indicates that not only do the phage carried genes successfully undergo transcription and translation but also that the resultant transferase enzyme is capable of functioning to some extent in the human galactose metabolic pathway.

PERSISTENCE OF PROKARYOTIC GENES IN EUKARYOTES

Galactosemic human cells infected with λpgal ($\text{K}^+\text{T}^+\text{E}^+$) survive longer than control cells (cells non-virus infected or infected with λpgal ($\text{K}^+\text{T}^-\text{E}^+$) or wild-type λ virus) in medium containing galactose as sole hexose source. The control galactosemic cells begin to show morphological evidence of metabolic stress (granular cytoplasm, rounding up, and detachment from the surface of the flask) at the end of 72 hours incubation in such a medium. However, cells used in the $^{14}\text{CO}_2$ production studies and infected with λpgal^+ have survived from 2 to 3 weeks longer than control cells when both were changed at the plateau phase of CO_2 release to a medium containing 0.1% galactose and dialyzed foetal calf serum. Nevertheless, we have been unsuccessful to date in isolating cells which, after exposure to λpgal^+ survive beyond five weeks in medium containing galactose as the sole hexose source. Normal cells (Led-130), on the other hand, have survived in

this medium for at least six weeks. However, the bacterial transferase enzyme and λ specific RNA have been detected 40 days after the initial infection of human galactosemic cells (Merril et al., 1971; Geier and Merril, 1972). In contrast, studies using bacterial viruses on plant tissue-culture cells have yielded longer survival on selective media (Doy et al., 1973c). Further understanding of the mechanisms of persistence in these systems may be greatly aided by genetic analysis.

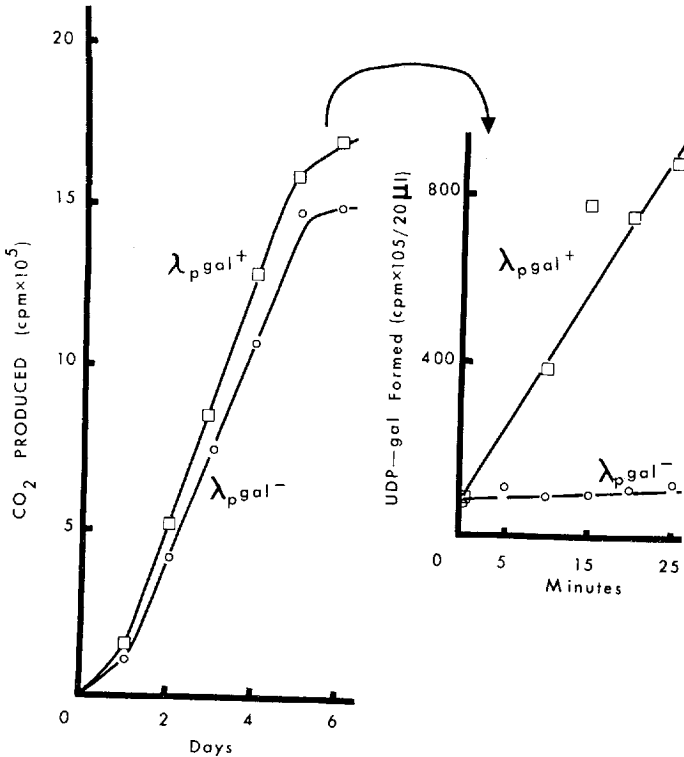


Figure 4. Transferase activity detected at the time of galactose metabolic enhancement in galactosemic fibroblasts. Cultures containing approximately 1×10^5 cells per flask in the logarithmic phase of growth were rinsed twice with Tris-buffered saline and inoculated with 1.0 ml of fluid containing bacteriophage (1.4×10^{10} p.f.u./cc) in Earle's BSS without sodium bicarbonate. After 1 hour at room temperature with frequent gentle rocking, the cultures were rinsed twice with Tris-buffered saline and fresh, hexose-free, medium MAB 87/3 containing penicillin and 10% dialyzed foetal calf serum was added. Galactose-1-¹⁴C (1.05×10^7 cpm; 98 mM) was added to each flask and silicone stoppers were inserted. The flasks were then incubated at 37°C. ¹⁴CO₂ was monitored as described for Figure 1. Different cultures were assayed after the 4th, 5th and 6th day of incubation for transferase activity.

TRANSDUCING VIRUSES AS BIOLOGICAL PROBES OF EUKARYOTIC CELL FUNCTION

A specially constructed transducing phage has already been used as a probe in the search for suppressors in animal cells and plant cells. The inability of λ_{pgal} ($K^{+}T^{-}E^{+}$) virus (constructed by M. Gottesman), which has an amber point mutation in the transferase gene, to induce transferase activity in animal cells (Merrill et al., 1971) and later in plant cells (Doy et al., 1973a) suggested that amber suppressors may not be widespread in eukaryotic cells. In fact, the introduction of a suppressor t-RNA as a result of infection of a transducing phage carrying a suppressor gene was shown to be lethal to plant cells (Doy et al., 1973a). Additional support for this view comes from the experiments of Schreier et al. (1973) using RNA from the *amb*₂ mutant of phage RL7 and a mammalian cell-free system which was shown to contain no detectable suppressor activity. Similarly, infection of plant protoplasts with an amber mutant of phage T3 failed to express the mutant phage function (Carlson, 1973).

VIROGENIC APPLICATION TO WHOLE ANIMALS

In an effort to develop an experimental approach to gene therapy in animals, λ viruses were administered to non-immune germ-free animals (Geier et al., 1973). Germ-free animals were chosen to minimize any potential interference of bacteria and bacterial viruses from the gastrointestinal tract. It was found that bacterial viruses are cleared from the body of the germ-free animal in the same exponential manner as any foreign colloidal particle. More than 99% of the injected virus is phagocytosed by the liver Kupffer cells within 30 min after infection. This rapid elimination can be slowed down by overwhelming the reticuloendothelial system with inert colloidal particles such as thorotrast (Kluge and Hovis, 1969; Brunner et al., 1960; Geier et al., unpublished data). All organs studied had similar exponential decreases in virus titre with the exception of the spleen, which retained large quantities of intact virus, capable of forming plaques, up to seven days after infection. These observations indicate that spleen is capable of non-destructive capture of antigens. This capture may utilize a non-phagocytic mechanism, described by Nagy et al. (1973). These studies with phage and germ-free mice serve as a basis for attempts to introduce specific genes carried by phages into normal laboratory mice. Unfortunately, mice which are defective in their galactose metabolism, to be used as a comparative experimental system alongside the tissue culture cells, are unavailable. However, the recent report on an almost complete lack of transferase activity in grey kangaroos may make this animal a more useful model system for the study of galactosemia in man (Stephens et al., 1974). Our initial experiments to pursue this possibility are now underway.

CONCLUSIONS

There is considerable evidence that prokaryotic genetic material does function in eukaryotes. These observations raise the possibility that the powerful tools developed to study the molecular biology of the specialised bacterial transducing viruses may be used to probe and alter higher organisms. Such application has already begun in the use of a bacterial virus containing an *amber* mutation to search for suppressor genes in the cells of higher organisms. Furthermore, it should be possible to use the techniques of producing new transducing phage (Shimada et al., 1972) to introduce a wide variety of *E. coli* genes into eukaryotic cells and use them as additional probes of the controls of transcription. When studies of the effects of λ transducing phage in cells in culture and whole animal systems have been completed, it may become possible to use λ as a vector for the introduction of *E. coli* genes into humans as an attempt to correct errors of metabolism. Alternatively, an attempt to isolate the human galactose genes using the procedures recently developed by Cohen et al. (1973) could be made by trying to select bacterial plasmids carrying the human galactose genes in *E. coli* cells deleted of their galactose transferase activity. Human DNA from such a plasmid could be first retested for its galactose gene activity in human galactosemic fibroblasts before any further consideration on the advisability of its introduction into patients.

ACKNOWLEDGMENTS

The authors would like to thank Dr G.L. Cantoni for his support and for making it possible for B.G.R. to visit and work in his laboratories.

REFERENCES

- Brunner, K.T., D. Hurez, R.T. McCluskey and B. Benacerraf. 1960. Blood clearance of ^{32}P -labelled vesicular stomatitis and Newcastle disease viruses by the reticuloendothelial system in mice. *J. Immunol.* 85: 99.
- Carlson, P.S. 1973. The use of protoplasts for genetic research. *Proc. Nat. Acad. Sci. USA* 70: 598.
- Cohen, S.N., A.C.Y. Chang, H.W. Boyer and R.B. Helling. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Nat. Acad. Sci. USA* 70: 3240.
- Cuatrecasas, P. and S. Segal. 1966. Galactose conversion to D-xylulose: An alternate route of galactose metabolism. *Science* 153: 549.

- Doy, C.H., P.M. Gresshoff and B.G. Rolfe. 1973a. Transgenesis of bacterial genes from *Escherichia coli* to cultures of haploid *Lycopersicon esculentum* and haploid *Arabidopsis thaliana* plant cells. In *The Biochemistry of Gene Expression in Higher Organisms* (J.K. Pollack and J. Wilson Lee, ed.) p.21. Australia and New Zealand Book Co. Sydney.
- _____, _____, _____. 1973b. Biological and molecular evidence for the transgenesis of genes from bacteria to plant cells. *Proc. Nat. Acad. Sci. USA* 70: 723.
- _____, _____, _____. 1973c. Time-course of phenotypic expression of *Escherichia coli* gene Z following transgenesis in haploid *Lycopersicon esculentum* cells. *Nature New Biol.* 244: 90.
- Geier, M.R. and C.R. Merril. 1972. Lambda phage transcription in human fibroblasts. *Virology* 47: 638.
- Geier, M.R., M.E. Trigg and C.R. Merril. 1973. Fate of bacteriophage lambda in non-immune germ-free mice. *Nature* 246: 221.
- Hotchkiss, R.D. 1954. Cyclical behaviour in Pneumococcal growth and transformability occasioned by environmental changes. *Proc. Nat. Acad. Sci. USA* 40: 49.
- Inouye, T., M. Tannenbaum and D.Y. Hsia. 1962. Identification of galactose-6-phosphate in galactosaemic erythrocytes. *Nature* 193: 67.
- Isselbacher, K.J. 1957. Evidence for an accessory pathway of galactose metabolism in mammalian liver. *Science* 126: 652.
- Johnson, C.B., D. Grierson and H. Smith. 1973. Expression of λ plac5 DNA in cultured cells of a higher plant. *Nature New Biol.* 244: 105.
- Kalckar, H.M., K. Kurahashi and E. Jordan. 1959. Hereditary defects in galactose metabolism in *Escherichia coli* mutants. I. Determination of enzyme activities. *Proc. Nat. Acad. Sci. USA* 45: 1776.
- Kluge, T. and T. Hovis. 1969. Ultrastructural localization of thorotrast in the reticuloendothelial system. *Amer. J. Path.* 54: 355.
- Merril, C.R. 1974. Transduction in mammalian cells. *Birth Defects*. In press.
- Merril, C.R., M.R. Geier and J.C. Petricciani. 1971. Bacterial virus gene expression in human cells. *Nature* 233: 398.
- _____, _____, _____. 1972. Bacterial gene expression in mammalian cells. *Advances in the Biosciences* 8: 329.
- Nadler, H.L., C.M. Chacko and M. Rachmeler. 1970. Interallelic complementation in hybrid cells derived from human diploid strains deficient in galactose-1-phosphate uridyl transferase activity. *Proc. Nat. Acad. Sci. USA* 67: 976.

- Nagy, Z., E. Horvath and Z. Urban. 1973. Antigen capture in chicken spleen during the primary immune response and relationship to phagocytic activity. *Nature New Biol.* 242: 241.
- Petricciani, J.C., M.K. Binder, C.R. Merrill and M.R. Geier. 1971. Galactose utilization in galactosemia. *Science* 175: 1368.
- Rogers, S. and P. Pfuderer. 1968. Use of viruses as carriers of added genetic information. *Nature* 219: 749.
- Russell, J.D. 1968. Improved transferase assay for cultured fibroblasts. *Biochem. Genetics* 1: 301.
- Schreier, M.H., T. Staehelin, R.F. Gesteland and P.F. Spahr. 1973. Translation of bacteriophage R17 and Q β RNA in a mammalian cell-free system. *J. Mol. Biol.* 75: 575.
- Segal, S., A. Blair and Y.J. Topper. 1962. Oxidation of carbon-14 labelled galactose by subjects with congenital galactosemia. *Science* 136: 150.
- Segal, S., A. Blair and H. Roth. 1965. The metabolism of galactose by patients with congenital galactosemia. *Amer. J. Med.* 38: 62.
- Shimada, K., R. Weisberg and M.E. Gottesman. 1972. Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. *J. Mol. Biol.* 63: 483.
- Stephens, T., S. Irvine, P. Mutton, J.D. Guppa and J.D. Harley. 1974. Deficiency of two enzymes of galactose metabolism in kangaroos. *Nature* 248: 524.
- Tedesco, T.A. and W.J. Mellman. 1971. Galactosemia: Evidence for a structural gene mutation. *Science* 172: 727.
- Thilly, W.G., H.A. Lozeron and W. Szybalski. 1973. Attempts to transform mammalian cells with DNA of phage ϕ 80dsuIII carrying the thymidine kinase gene. In Possible Episomes in Eukaryotes. (L.G. Silvestri, ed.) p.250. North-Holland Publishing Co., Amsterdam.
- Topper, Y.J., L. Laster and S. Segal. 1962. Galactose metabolism: Phenotypic differences among tissues of a patient with congenital galactosaemia. *Nature* 196: 1006.

Resumé of the conference

473

Dan L. Lindsley

CHROMOSOMAL MOLECULES

DNA I begin this resumé by first recounting something about the molecules that make up the chromosome and then discussing my impression of how these molecules might be arranged within the chromosome to suit its functional needs. Throughout this discussion I propose to speak in terms of kilobases, i.e., 10^3 base pairs in double-stranded or 10^3 bases in single-stranded DNA. As a geneticist I find myself much more comfortable thinking in terms of these units of length rather than in terms of either microns or daltons.

In 1957 Taylor, Woods and Hughes showed autoradiographically that the chromatid is composed of two subunits which replicate and segregate as units and which extend from one end of the chromatid to the other. Later Taylor (1958), by investigating the ways in which exchanges between sister chromatids are restricted, was able to infer that the two chromatid subunits are unlike. It remained for Brewen and Peacock (1968) to show, by studying the restrictions in the way that proximal sister union of isochromatid breaks may occur, that the difference between the two strands is one of polarity. All of these observations were consistent with the hypothesis that the chromatid is composed of a single DNA molecule; however, cytologists and geneticists were not yet prepared to think in terms of molecules of the colossal dimensions demanded by such a hypothesis. The more recent work with *Drosophila* (Kavenoff and Zimm, 1973) and yeast (Petes and Fangman, 1972; Cryer et al., 1974) and work described by Fangman in this symposium have provided strong evidence for the supposition that in fact a single DNA molecule extends from one end of a chromatid to the other. However, this molecule may not be isopolar throughout its length. We (Peacock et al., 1973) have published data showing that newly induced ring chromosomes replicate to produce double-sized dimer rings more frequently than

can be accounted for by sister chromatid exchange; the observations can be explained by postulating occasional switches in polarity of the DNA molecule (i.e., one 3'-3'/5'-5' bond per 10μ of metaphase length).

Some perplexing results were presented here by Ford that can be most easily interpreted in terms of two parallel DNA molecules per chromatid in cells of human amniotic fluid. It is very hard to reconcile her observations with those cited above which lead to the unineme hypothesis. Reaction to Ford's data was one of consternation, and there was considerable discussion as to where the error could be; most conferees found it less troublesome to suppose that an error had been made than to change their habits of thinking. I consider it of utmost importance to examine the segregation of label in these cells so as to determine whether by this criterion too they fail to conform to conventional expectation.

The organization of chromosomal DNA is encompassed by its base sequence, and its secrets must be sought in the sequential diversity within this tremendously long molecule. Until recently the methods that have been available for the investigation of sequence diversity have been somewhat limited. Chromosomal DNA may be sheared to pieces of a predetermined length, of extremely small size, relative to the whole molecule, and then fractionated according to physical chemical properties, such as buoyant density or reassociation kinetics.

Satellite DNAs may be isolated either on the basis of having a buoyant density different from that of the remaining fragments or on the basis of their extremely rapid reannealing. These fractions are composed of arrays of simple repeating nucleotide sequences. We heard a detailed analysis of the satellite DNAs of *Drosophila melanogaster* from the Canberra group of Peacock, Brutlag, and Goldring. They have isolated four satellite fractions on the basis of their characteristic buoyant densities and have made considerable progress in determining their sequences. Furthermore, their evidence suggests that each satellite of homogeneous buoyant density is composed of a family of isomeric variants of the simple repeating sequence and that the different isomers occur on different shear segments rather than being interspersed. The evidence further suggests that the satellite sequences are arranged in long arrays (>600 kb) of one simple sequence in tandem with long arrays of other simple sequences. *In situ* hybridization of transcripts from the purified satellites shows them to be localized to the centromere regions of all the chromosomes. Furthermore, the fraction of the total DNA that can be isolated as satellite is equivalent to the length of chromosome that is occupied by constitutive

heterochromatin which is localized around the centromeres. There is thus a strong possibility of a 1:1 relation between constitutive heterochromatin and DNA. Of the four satellite families one appears to be confined to chromosome 2 and the Y chromosome whereas the others are apparently carried on all chromosomes, though the intriguing possibility exists that different isomers will be found to be chromosome specific. One further provocative observation is that sequences of each of the four satellite DNAs exist in the euchromatic band 21D1-2 at the tip of the left arm of the second chromosome; the meaning of this fact is obscure, but the observation provides a fruitful area for investigation. An interesting topic for investigation in the future is the relation of these satellite sequences to the known properties of constitutive heterochromatin. Why, for instance, do euchromatic genes when brought into juxtaposition with pericentric heterochromatin frequently variegate?

A second type of DNA fraction which may be identified within the population of sheared fragments is that containing moderately repeated sequences. Combining studies on the kinetics of hybridization and electron microscope observations of reannealed molecules, Davidson has determined that in the sea urchin this material occurs in short sequences averaging about 0.35 kb in length and separated from one another by stretches of non repetitive DNA from 0.5 to 4 kb long. Approximately 80% of the non-satellite DNA is organized in this manner, whereas the remaining 20% appears to be free of repeating sequences; this partition may be a function of the lengths of fragments studied, and the 20% may merely show slightly looser interspersion. Davidson has observed similar degrees of interspersion for a wide array of organisms. Nevertheless, preliminary indications are that in *Drosophila* the interspersion is approximately one order of magnitude coarser; however, I am unable to judge whether this difference is due to procedure or is a species characteristic.

A second method for examining the sequential organization of chromosomal DNA makes use of RNA molecules transcribed therefrom. These are the heterodisperse population of molecules found in nuclei and known by the rather colorful name of heterogeneous nuclear RNA (HnRNA). These molecules have been isolated in the size range from 15 to 45 kb, and Darnell has made initial progress in identifying some of the features common to this heterogeneous collection of molecules. Briefly, the 3' end contains messenger RNA which is transported to the cytoplasm after processing, and the 5' end, which is degraded during processing of the molecule, contains short complementary segments which self-anneal as well as oligo (U)-containing regions. According to Davidson, HnRNA also contains transcripts of

the moderately repeated sequences discussed in the previous paragraph. The 3' terminus of most mRNA molecules consists of a polyadenylated chain of about 200 residues, most of which are added after transcription. However, Lodish et al. (1974) have shown in *Dyctiostelium* that the first 25 or so actually exist in the primary transcript.

One of the exciting aspects of the conference was Hogness's description of a newly developed method for randomly selecting a sheared DNA fragment and cloning it. This is accomplished by inserting it into a bacterial plasmid and growing it to high titre in bacterial culture. The primary advantage of this approach is that it allows the investigation of a population of sequentially identical molecules. *In situ* hybridization of such cloned DNA molecules to polytene chromosome preparations is effective in identifying their site of origin in the chromosomal complement. Hogness described three plasmids containing *Drosophila* DNA. They contain segments of *Drosophila* DNA of 3, 8, and 13 kb which on the basis of reannealing kinetics are represented in the genome in 87, 4, and 2 copies respectively. The 87 copies localize to fourteen places in the polytene complement and the other two localize to a single site. Hogness and his colleagues are currently attempting to insert into plasmids segments of DNA that are larger (50 to 100 kb) than the average chromomere (26 kb). Once isolated, the foreign DNA in these plasmids may be further segmented by means of restriction endonuclease and mapped in even finer detail. As the library of such plasmids expands, this approach promises to provide important new insights into chromosome and chromomere organization.

Stepping up an order of magnitude, Fangman described the potential for the isolation of the intact DNA molecules of individual chromosomes of yeast, the smallest of which contains approximately 150 kb. If this can be achieved it should provide a promising system for the investigation of the sequential organization of an entire chromosome including centromere determinants and chromosome ends.

During my own sabbatical sojourn in Canberra last year I attempted a conceptually similar approach. It is possible to produce centric derivatives of the X chromosome that are smaller than chromosome 4 and place them in a stock in which they are the only small elements in the complement. Such minisomes contain all the elements of higher eukaryotic chromosomes including termini and a centromere, as do yeast chromosomes. In addition they contain both a small euchromatic region organized into chromomeres and a larger satellite-containing heterochromatic segment. If these minisomes or their DNA molecules can be isolated they will provide

material enriched for X chromosome satellite sequences as well as end and centromere determinants. However, the size of the smallest minisome achieved (ca. 1,500 kb) is an order of magnitude larger than that of the smallest yeast chromosomes, and its isolation remains complicated from a procedural standpoint.

Proteins Turning now from chromosomal DNA, let me dwell on chromosomal proteins for a moment. Kornberg told us a stunning story of the current ideas on the relation between histones and DNA being developed by the Cambridge group. Histones F2A1, F2A2, F2B, and F3 are present in equimolar amounts. F2A1 and F3 always occur associated in an $\alpha_2\beta_2$ type of tetramer; F2A2 and F2B show evidence of multimeric association. X-ray diffraction studies of chromatin reveal a repeating structure of 100Å diameter. Kornberg adduces evidence suggesting that each of these structures contains one F2A1₂F3₂ tetramer plus two molecules each of F2A2 and F2B, and 0.2 kb of DNA. He also presented the suggestion that the links between these 100Å units are flexible, resulting in a structure analogous to a string of beads. This picture, arrived at on the basis of physical and chemical considerations, is surprisingly concordant with electron micrographs of chromatin published by Olins and Olins (1974) and Woodcock (1973) earlier this year except that the dimensions of the beads on their strings are 70Å. It appears that DNA in the linking region is more susceptible to endonuclease than the rest inasmuch as 200 nucleotide pieces are released upon digestion of chromatin with certain nucleases. This picture provides appealing stoichiometry between four of the five major histones (there are half as many molecules of F1 as the others but it does not contribute to the repeating unit) and DNA. No such information was presented for the way in which nonhistone proteins contribute to chromosome organization. Hill presented evidence that the metabolically extremely active lampbrush chromosomes of the newt oocyte become greatly enriched for two nonhistone proteins; these proteins presumably function at the time of transcription and although associated with the chromosomes do not form a fundamental part of its fabric.

FUNCTIONAL ORGANIZATION OF CHROMOSOMES

I would now like to shift the discussion from the nature of molecules present in the chromosomes and their relations to one another to a consideration of how the chromosome is organized to perform the various functions characteristic of chromosomes. Chromosomal functions that I will consider briefly are the following: transcription, replication, separation of replicas, condensation, and movement, all of which are necessary for the completion of every cell division as well as synapsis and recombination which are confined to the first meiotic division.

Transcription In the bacterial operon the DNA sequence to which the DNA-dependent RNA polymerase binds, the promoter, may be separated from the structural gene by an operator sequence which functions in the control of transcription, and although transcribed is not translated. A major difference between prokaryotes and eukaryotes appears to lie in the levels of regulation necessary to control time, extent, and place of gene action. The number of such levels in eukaryotes is large compared to the relatively very simple control systems in the prokaryote. This being so, it is tempting to consider the extremely long processed region of eukaryotic HnRNA as having been transcribed from the DNA that is the eukaryotic counterpart of the operator sequence of the bacterial system. If mutation within this hypothetical regulatory sequence were capable of interfering with the control of transcription, then many induced mutants might be expected to have polar effects and to fail to complement other such alleles.

An interesting observation in this connection is that of Abrahamson et al. (1974) that although the rate of induced mutation of specific loci in different organisms varies over three orders of magnitude, the specific locus mutation rate divided by the amount of DNA in the haploid complement is virtually a constant, values varying over but a three-fold range. One possible explanation that they propose for this increase in target size as the size of the genome increases is that gene size is correlated with genome size. An additional observation that is consistent with the same interpretation has been presented by Sandler and Hecht (1973). They note that the fraction of the genome that can be duplicated or heterozygously deleted and yet be compatible with survival are comparable in man and *Drosophila*, even though the quantity of DNA comprising these fractions is an order of magnitude greater in man.

Adhering to the view that the eukaryotic genome contains long transcribable units in which the 3' end of the transcribed strand contains regulatory sequences and the 5' end contains one or more structural genes, we might ask what are the consequences of breaking and rearranging the eukaryotic chromosome? In the data presented in Table 1, the recessive lethal frequencies are given for a series of rearranged chromosomes compared with a sample of non-rearranged chromosomes from the same experiment; the difference represents the incidence of rearrangement-associated lethals from which the probability that a breakpoint is associated with a lethal may be computed. The surprising observation to me is that so large a fraction of disruptions of the chromosomal DNA has no apparent effect on the well-being of the organism. For those who are content with the idea that most of the DNA is non-informational, on the other hand, the observation that 10 to 40% of the breaks are lethal

is probably surprisingly high. A break that occurs within a transcriptional unit is very likely to separate the structural gene from the point of transcription initiation. Given two such breaks in a translocation event two new promoter-structural gene combinations could arise that are abnormally regulated. Alternatively, a ruptured transcriptional unit could be translocated to a non-transcribed region, thus inactivating the structural gene. Also both breaks could be in non-transcribed regions of the genome. Structural genes in the *cis* configuration with an abnormal constellation of regulatory sequences might continue to function but in an abnormal quantity, time, or place and have a dominant effect, either lethal or visible. It is interesting to note that a large fraction of dominant visible mutations in *D. melanogaster* are associated with chromosome rearrangements. Dominant visible mutations are quite rare, but dominant lethals are a common consequence of X-irradiation, and it may be that an appreciable fraction of chromosome aberrations are dominant lethals and thus never scored.

If one makes the questionable assumptions that translocations with both breaks in transcribed regions are dominant lethal, that both breaks in non-transcribed regions are viable, and that one in each are recessive lethal, then a simple minded computation from Ives' (1972) data leads to an estimate that one-fourth of the breaks, and inferentially one-fourth of the DNA, are in transcribed regions: correcting for 25% non-transcribed satellite DNA yields a 2:1 ratio of non-transcribed to transcribed DNA in the euchromatin of the X, in remarkable agreement with Laird's (1973) estimate that one-third of the unique sequences are transcribed. Concluding from the data presented by Judd that the chromomere (30 kb) is the unit of transcription we can crudely estimate it to be composed of 20 kb not transcribed and 10 kb transcribed DNA; the latter value could represent the expected size of *Drosophila* HnRNA. These considerations assume that rearrangement breakpoints are randomly distributed in the genome, but there are some indications that this may not be so. For example, among the 48 recovered translocations between the X and the Y chromosome, Merriam (personal communication) found five with identical X chromosome breakpoints. He inferred the identity from the fact that all X chromosome combinations of the distal segment from one translocation with the proximal segment from another were fully viable. Such a relationship indicates that there is no transcribed material between any pair of the breakpoints tested. The probability that five breaks out of 48 in the 33,000 kb X chromosome would be so close to each other on the basis of random breakage and rejoining seems remote indeed.

TABLE 1. Estimation of the fraction of rearrangement breakpoints that are also lethal from the relative frequencies of recessive lethals among rearranged and non-rearranged chromosomes from the same experiment

Irradiation Stage	Dose (kr)	Rearrange-ment scored	Chromo-some tested for recessive lethal	P(1 R)	P(1 R̄)	n	Probabil-ity* that a break-point is also lethal	Reference
Mature sperm	4	T(X;A) + T(Xe;A) + T(Xh;A)	X X X	0.383 0.437 0.250	0.130 0.130 0.130	1 1 1	0.291 0.353 0.138	Williams (1965)
Mature sperm	4	T(X;A)	X	0.205	0.068	1	0.147	Wagoner (1965)
Mature sperm	1	T(2;3)	2+3	0.554	0.135	2	0.282	Ives (1972)
2	T(2;3)	2+3	0.646	0.223	2	0.325		
Spermatids	1	T(2;3)	2+3	0.631	0.228	2	0.313	
Mature sperm	2	In(1)	X	0.588	0.139	2	0.308	Lefevre (1974)
Spermatids	1	In(1)	X	0.584	0.128	2	0.310	
2	In(1)	X	0.731	0.242	2	0.404		
Mature sperm	3	T(X;A)+In(1)	X	0.304	0.063	1-2	0.259-0.138	Lefevre (1974)
Immature sperm	3	T(X;A)+In(1)	X	0.316	0.037	1-2	0.274-0.148	
Spermatids	2	T(X;A)+In(1)	X	0.378	0.119	1-2	0.294-0.160	

*The probability that a breakpoint is also lethal is estimated from the expression $1 - [P(\bar{1}|R) / P(\bar{1}|\bar{R})]^n$ where $P(\bar{1}|R)$ is the probability of no recessive lethal given a rearranged chromosome, $P(\bar{1}|\bar{R})$ is the probability of no recessive lethal given a non-rearranged chromosome, and n represents the number of breakpoints assumed to be tested for recessive lethality.

†Xe and Xh refer to the euchromatic and heterochromatic portions of the X chromosome.

This fanciful speculation on a polar arrangement of regulatory and structural information should be treated cautiously. Chovnick and his colleagues (Gelbart et al., 1974) have failed to find any evidence for mutations of a regulatory nature clustered at one end of the map of the *rosy* locus. They find that null alleles (some of which could be regulatory) and electrophoretic alleles (none of which can be regulatory) are interspersed within the map of the locus.

In summary, although the size of transcriptional units may vary among organisms, they appear to involve tens of kilobases and be of the same order of magnitude as chromomeres. The next five years or so should witness considerable elucidation of their structural and functional organization, by virtue of studies as diverse as HnRNA behavior, replicating eukaryotic DNA plasmid complexes, and high resolution cytogenetic analysis.

Replication We are concerned with at least two levels of chromosome organization with respect to replication. The finest level at which DNA replication is discontinuous is represented by the 0.2 kb Okazaki fragment. It is not clear what signals the initiation of successive Okazaki fragments, but it seems unlikely that DNA sequence is important. If it were, sequential signals would require such dense distribution as to interrupt individual structural genes. In line with such reasoning Martin reported a lack of any regularity in the RNA sequences that initiate Okazaki fragment synthesis in polyoma virus.

Periodic sites of initiation of DNA synthesis are probably incorporated into the second level of organization for chromosome replication. That these must exist has been obvious for a long time from the observation that even the most minute chromosome fragments induced by X-irradiation are capable of replicating in synchrony with the remainder of the chromosome complement. Moreover, quantitative consideration of the rate of DNA synthesis and of the time required to replicate a chromosome complement demand multiple initiation sites. Hogness has shown us that in the rapidly cleaving nuclei of the *Drosophila* zygote (cycle time = $9\frac{1}{2}$ minutes, S-period < 4 minutes) bidirectional replication starts at initiation sites that are separated by 3.4 kb or multiples thereof. Fangman presented data showing that nitrosoguanidine may cause large increases in the mutation rate of particular genes depending on the portion of the S-period at which it is administered. Since nitrosoguanidine is thought to act preferentially at points of replication, this observation suggests that specific segments of DNA replicate in a characteristic temporal sequence and that initiation sites for the DNA segments have defined positions on the genetic map.

Initiation of DNA replication appears to be controlled in two ways. One mechanism determines whether an individual initiation site will act as such at any time during a particular synthetic period and a second mechanism determines when during the S-period initiation is triggered. An example of the first type of control was presented by Hogness who reported that the eye forms or replication bubbles are much more densely distributed in cleavage nuclei where the S-period is 3-4 minutes than in cultured cells where the S-period lasts 6 hours. He postulates that initiation sites cannot operate in compacted chromosome regions so that the functioning of an initiation site during a particular cycle is controlled by the mechanism governing chromosome condensation. Graves described control of a different type as observed in the asynchronous initiation of replication in different parts of mouse and hamster chromosome complements. In progeny of mouse-hamster cell hybrids which had lost varying numbers of hamster chromosomes, replication started synchronously in the chromosomes derived from both species. Irrespective of the chromosome constitution of the hybrid, hamster chromosomes maintained their characteristically shorter S-period. Moreover, chromosome segments which were late labeling in diploid hamster cells were also late labeling in the hybrid cells. She surmised that a sequence of replication is programmed for each chromosome and that once the S-period starts, the program operates independently of extra-chromosomal signals.

Generally speaking, the DNA complement is replicated once and only once during each S-phase so that parity is maintained among various chromosomal regions. Two cases of departure from this general rule deserve special mention: amplification and under-replication. Gall described the extra-chromosomal replication of ribosomal DNA, but as he indicated, the interesting question with respect to eukaryotic chromosome organization is how the rDNA segment is first disengaged from the chromosome so as to serve as template for extra-chromosomal amplification. He offered no insights, but in describing his observations of amplification in *Tetrahymena* he pointed to the possibility of synchronizing the initiation of amplification by regulating conjugation. Such experimental control of initiation would provide a model system for studying extra-chromosomal replication. If so, we can be optimistic that the next few years will provide at least some answers to this vexing question.

The reciprocal of extra-chromosomal amplification is the under-replication of satellite sequences during polytenization when all but a small fraction of the genome is amplified. The resulting situation in which two arms of a chromosome are amplified a thousand-fold but are connected to each other through a relatively complicated centromere region poses difficult problems for resolving

the structure of polytene chromosomes. It might be thought that under-replication of the centromere region would render it fragile from the standpoint of chromosome continuity. Yet, when the aggregated centromeres at the chromocenter are disrupted by squashing, it is generally observed that the association between the two arms of a chromosome persists even though homologous chromosomes may be separated from each other. The topological relations among strands as they traverse the centromere constitute an intriguing question for several investigators, but it is one for which an experimental approach still needs to be fashioned.

Separation of sister chromatids Although seldom mentioned, and nearly always ignored in diagrammatic representations, replication of a double helix should result in two relationally coiled double helices in which the period of the relational coil is equal to that of the DNA double helix. Intertwined strands of such tremendous length should face insurmountable difficulties in disengaging from one another. Chromosomes apparently solve this problem coincidentally with replication. As can be seen from the electron micrographs presented by Hogness, the eye forms in the replicating DNA of *Drosophila* cleavage nuclei deny the presence of any difficulty in separation of daughter replicas. Since their disengagement is a problem that arises at every twist of the double helix it must be solved by breakage and rejoining of at least one of the sugar-phosphate backbones. Such breakage could conceivably occur at each twist of the helix, or, if at wider intervals, each break would serve to resolve an accumulation of twists. I know of no information that bears on how a chromosome might be organized for replica disengagement. It might be sufficient to have a cleavage enzyme whose substrate is relationally coiled DNA molecules (Wang, 1971; Champoux and Dulbecco, 1972). Alternatively, the chromosome could house enzyme specific sites at which interlocks are resolved. These questions must be left to future investigations. However, it should be noted that the disengagement of replicas is not accompanied by a disengagement of sister chromatids. They remain in close parallel association in all chromosomes and even in the most minute acentric chromosome fragments as though some binding persists between them.

Condensation The DNA of chromosomes is probably never fully extended even in the interphase chromosome; it seems attractive to think of the structure described by Kornberg in which 0.2 kb segments are packaged in 100A units, or in a packing ratio of seven, as being the maximally extended state of DNA in the interphase chromosome. Superimposed on this is a higher order of organization, in which there is an intermittent condensation of the chromatin thread into chromomeres, which comprise 26 kb on the average in *Drosophila melanogaster*. These chromomeres give rise to bands in the polytene

chromosome, where hundreds of replicas of a chromosome strand are arranged in a parallel array with their chromomeres in register. The constancy of the banding pattern is evidence that compaction into chromomeres occurs according to some plan that is programmed into the nucleotide sequences. Exactly how proteins and DNA interact to achieve such discontinuous compaction is a matter of utmost interest to which molecular cytogeneticists should turn their attention. Perhaps the plasmids containing *Drosophila* DNA will provide tools for the investigation of this question. Discontinuous condensation along the chromosome length is evident at still another higher level of compaction. This is seen at the pachytene stage where the mapped chromomeres are found to be appreciably fewer and much larger than those present in polytene chromosomes. The constancy of pachytene pattern argues that the program for this higher degree of condensation is also contained within the nucleotide sequence. Extending this argument, one is persuaded to conclude that the information for the ultimate level of condensation seen in metaphase chromosomes must also be housed in the nucleotide sequence. Most probably the metaphase chromosome is compacted in a rigorously defined pattern that is repeated in every cell division cycle.

The organization that specifies how a chromosome will condense must be local and repeated at relatively short intervals. It is not clear what function is served by this characteristic, intermittent condensation, but its regularity is impressive. Although Crick (1971) has proposed a model of chromosome organization in which localized condensation plays a functional role, supporting evidence is very limited. For the present, it is worth emphasizing the evidence for local autonomy in regulating condensation. We note that polytene chromosome segments retain their characteristic banding pattern no matter how complexly they are rearranged, and the same holds for the chromomere pattern of rearranged pachytene chromosomes. Finally, the smallest X-ray induced acentric fragments that can be scored not only replicate but also condense in synchrony with the remainder of the complement.

The genetic control of chromosome condensation is another area of contemporary interest in which many questions have been raised but few answers given. The regions around the centromeres in *Drosophila*, which the Canberra group has shown to be largely composed of satellite DNA, were termed heterochromatin by the early cytologists because of their regular allocyclus during mitosis. These regions attain a high degree of compaction during prophase well in advance of the remainder of the chromosome complement. Why late replication should be correlated with early condensation is a

question on which I have seen no speculation. Furthermore, the reason why regions rich in satellite DNA should preferentially manifest allocyclic behavior is a challenge for future research.

Satellite-rich regions are referred to as constitutive heterochromatin because they are allocyclic in most mitotic prophase, although according to Hogness, not in those of *Drosophila* cleavage nuclei. Other chromosomes or their parts which may become allocyclic under certain conditions are termed facultative heterochromatin. They differ from constitutive heterochromatin in containing a representative array of DNA sequences rather than being confined to a specific fraction, the satellite DNA. Examples of facultative heterochromatin are the inactive X chromosomes of female mammals as discussed by Cooper, and the entire paternally-derived chromosome set in most somatic cells of certain coccids (Brown, 1966). Another example is the precocious condensation of the X chromosome in the primary spermatocytes of those species in which the male is the heterogametic sex, an event that may be important for the normal course of spermatogenesis (Lifschytz and Lindsley, 1972). Significantly, condensation is not precocious in either the first zygotic mitoses or in the premeiotic mitoses, and indeed, the studies described by White indicate that X chromosome inactivation may even be retarded in the last premeiotic mitoses. How one chromosome or a particular segment of a chromosome become programmed to condense out of phase with the remainder of the complement is a difficult question to approach. We feel, however, that X chromosome allocy, which is so important for normal spermatogenesis, offers a promising system for its investigation.

Movement In all but the first meiotic division daughter chromatids move apart at anaphase. Their movement is generally effected by spindle microtubules which attach to one localized site on each chromosome, the centromere. The evidence in *Drosophila* (Kavenoff and Zimm, 1973) suggests that the chromosomal DNA passes, without interruption of its continuity, through the centromere. This in turn suggests that the centromere region of the DNA contains sequences that bind specific proteins and thus confer the observed affinity for microtubular elements. Inquiry into the molecular organization of centromere regions has barely begun.

Synapsis We now turn from those chromosomal functions characteristic of all cell divisions to those which are peculiar to the first meiotic division, synapsis and recombination. At the beginning of synapsis homologous chromosomes become associated with one another through a specialized structure, the synaptonemal complex. Stern referred to this association as a rough alignment of homologues since the synaptonemal complex separates them by about 1500Å. Furthermore,

as Gillies' serial reconstructions of pachytene nuclei showed, the synaptonemal complex extends from one end of the bivalent to the other and its length is but a small fraction of that of the chromosomal DNA. This observation implies that DNA is intermittently associated with the synaptonemal complex. An excellent candidate for the sites of association is the DNA described by Stern as 'Zygotene-DNA', whose replication is deferred from premeiotic S until the initiation of synapsis at zygotene. This DNA appears to be single-, or at most, a few-copy DNA which one would presume to be a requisite for the specificity of recognition of homologous regions. Stern and Hotta's data indicate that zygotene DNA exists on the average of 3-4 kilobase segments distributed at thousand kilobase intervals. The relations between zygotene DNA, the synaptonemal complex, and homologous synapsis represent intriguing problems for future investigation.

Recombination Recombination has been an enigma to geneticists for many years; they have measured it and used it in mapping the genetic positions of mutant genes, but the information was accumulated without reference to events at the molecular level. More recently, these events have been vigorously pursued but in general, they have been elusive to the various experimental probes. Lefevre (1971) argues that over a major portion of the X chromosome in *Drosophila*, recombination within a segment is directly proportional to its DNA content. However, this proportionality breaks down at the ends of chromosomes and over large segments of the autosomes, where, for instance, the distal half of 3L is 40 units long and the proximal half is 8 units. Thus, on a gross level, there are regional differences in the amount of recombination per unit of DNA. Catcheside has argued that eukaryotic organisms have developed mechanisms for limiting the amount of recombination in accounting for the generally observed decrease in the rate of recombination as the amount of DNA increases. It may be that the control is regional, because all treatments that increase recombination in *Drosophila* lead to maps that are more nearly proportional to DNA content. One can argue at a much finer level that the chromosome is organized in a repetitive mode for effecting recombination. Observations on the polarity of gene conversion point to a process involving a chromosome segment of variable length with a defined starting point. The data indicate that such starting points are probably densely distributed along the DNA.

Catcheside presented a detailed analysis of the effects of three different dominant suppressors of crossing over (rec^+) in *Neurospora crassa* which reduce recombination or conversion at various places unrelated to their own locus. These dominants affect

segments about one crossover unit in length from which he argues that there is one site for the recognition of the *rec⁺* product per unit of recombination, or about 500 such sites in *N. crassa*. Dividing this number into the DNA content of the haploid complement of 5×10^4 kilobases yields an estimate of one such recognition site per 100 kilobases. Similar arguments suggest a similar dispersion of sites at which enzymes promoting recombination bind.

From a completely different aspect Stern reported to us about the repair synthesis of DNA that takes place during pachytene. This DNA is repetitive in nature, being of intermediate C_0t , and appears to represent the sites of pachytene-specific nicking by endonuclease. These sites are about 0.17 kb long and are nicked at 160 kilobase intervals. This is the average distance between nicks; however, we should not dismiss the possibility that nickable sites are actually more closely dispersed and the probability that an endonuclease-sensitive site will be nicked is less than one. The number of sites that are nicked is four orders of magnitude higher than the number of chiasmata (Hotta and Stern, 1974) observed in lily; so, although the involvement of pachytene nicking in recombination seems highly probable, it is certain that most nicks do not participate in recombination or conversion events. The problem that defies the imagination is how, in the bivalent, two homologous segments of a bivalent manage to find one another within the great mass of DNA. It seems logical, as Stern postulated, that the molecular pairing necessary for recombination must take place within the synaptonemal complex, but the amount of DNA involved is so small as to be incapable of visualization.

CONCLUSION

The thrust of this meeting seems to me to have been that every aspect of chromosome biology investigated rests on a periodicity in chromosome organization. Had these periodicities been comparable it would have been pleasing to think of the chromosome as composed of units containing one of each of the different functional segments. However, as can be seen from the figure, the periodicities for the various functions vary over several orders of magnitude. The size in kilobases of each type of repeating unit is represented on a logarithmic scale which extends from 10^{-2} as the size of the repeating units in satellite DNA up to 10^4 which is the size range of eukaryotic chromosomes. Within this range perhaps only the 10^{-1} level is unassociated with a periodicity in the base sequences, being more likely a periodicity attributable to protein structure. Among the remainder we see that successive orders of magnitude in the size of the repeating unit are associated with replication,

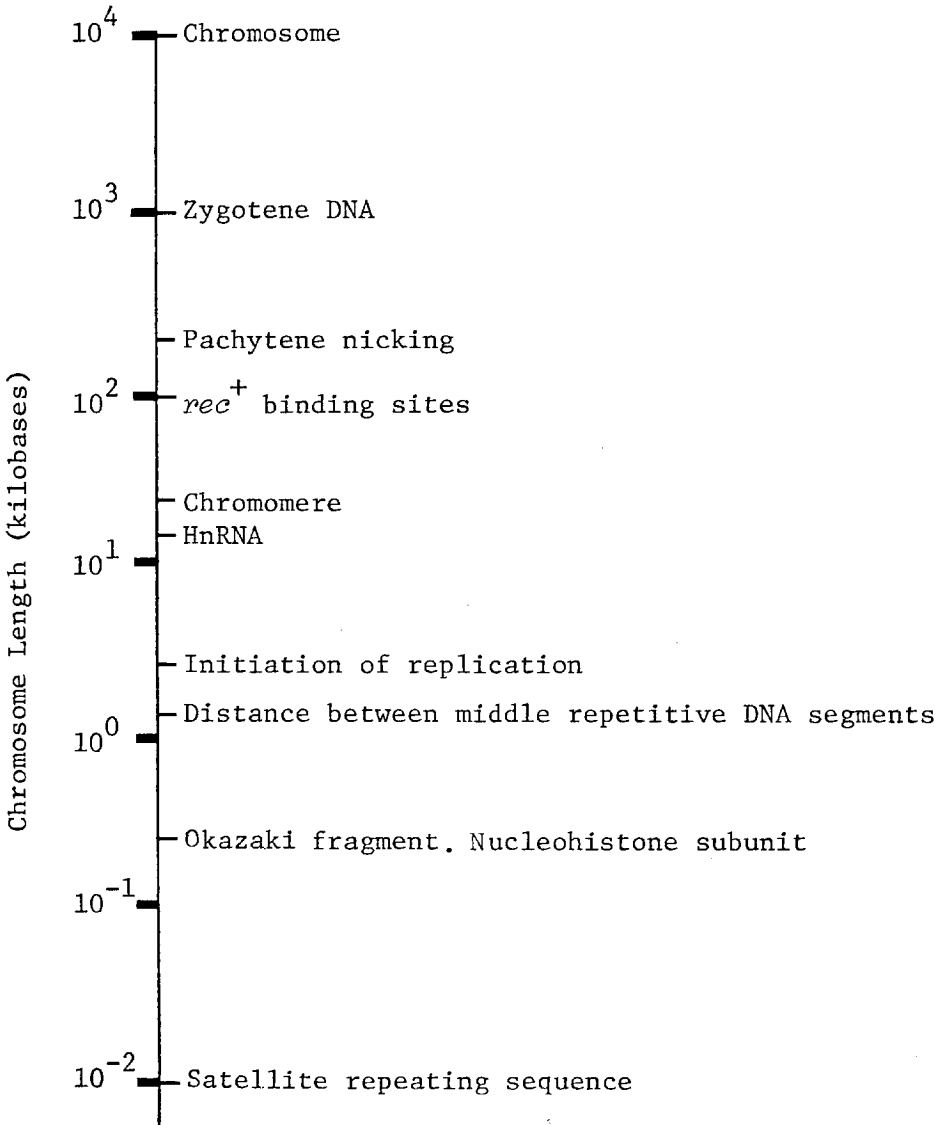


Figure 1. Estimated sizes of diverse repeating units of chromosome organization.

transcription, recombination and synapsis. One should be cautioned about taking these estimates too seriously as they represent means of distributions whose variances are unknown; a more serious objection to the data is that they are derived from the entire array of eukaryotic organisms, from yeast to man; the scale nevertheless presupposes an underlying constancy in the various periodicities among eukaryotic species, a supposition which, as yet, has little supportive evidence.

EPILOGUE

Now that I have outlined my impressions of the general thrust of this very successful conference, for which our hosts deserve special applause and thanks, I would like to discuss briefly some problems of chromosomal organization that I see as a *Drosophila* geneticist and which I believe were not considered during the conference.

The first question has to do with relating the satellite nature of constitutive heterochromatin with its unusual genetic properties. The genetic properties of the centromeric heterochromatin have been investigated only in *D. melanogaster*, but in that organism we have a wealth of information. The data on satellite sequences are consistent with the interspersion of sites of initiation at 3-4 kilobase intervals just as is found in the DNA from cleavage nuclei by Hogness and his colleagues. Hogness postulates that compaction of chromatin inactivates initiation sites and, the rate of fork migration being a constant, thereby lengthens the S-period. He further speculates that heterochromatic regions are more uniformly compacted at the time of S, and therefore have fewer initiation sites, their number determining the duration of the S-period. In this way, he manages to shift the problem of the temporal control of replication to the temporal control of condensation. To reconcile all the observations on chromosome replication with this model alone seems a bit contrived. We must suppose that in cultured *Drosophila* cells, centromeric heterochromatin, but no euchromatin, suffers a degree of compaction that is sufficient to block all DNA synthesis in centromeric regions at the initiation of S-phase. Some time after the initiation of S the heterochromatin must decondense to allow initiation of DNA synthesis. If, as occurs in the grasshopper according to White, the heterochromatin replicates more rapidly than the euchromatin, the density of initiation sites exposed in heterochromatin must be greater than that originally exposed in the euchromatin. The heterochromatin must then recondense at the end of S at such a rate as to overtake the euchromatin and achieve virtually complete compaction by prophase to achieve the observed allocyclus. By contrast, in embryonic nuclei the entire complement must begin S in the completely extended form to be compatible with its extremely short duration, a process that correlates with the apparent absence of allocyclus in these nuclei. How these complicated temporal sequences are controlled is anybody's guess. One can imagine that repetitive sequences of satellite DNA provide an organization capable of rapid condensation. It is difficult, however, to ascribe a function to the allocyclic completion of condensation by the centromeric heterochromatin.

Another characteristic of centromeric heterochromatin is the absence of meiotic (but not mitotic) recombination. This may suggest the absence of sequences necessary to initiate exchange interspersed among the satellite sequences. If so it suggests that mitotic crossing over is not dependent upon such sequences.

Next I would like to consider the generation of rearrangements, with special reference to the participation of heterochromatin in this process. Early workers (Bauer et al., 1937) found that when mature sperm of *Drosophila* are irradiated, the breakpoints of the resulting aberrations, as determined by polytene chromosome analysis, are distributed among chromosome arms and parts of arms roughly in proportion to their metaphase lengths and, inferentially, according to their DNA content. A more detailed analysis by Kaufmann (1946), however, on the distribution of 1,400 X-ray induced X chromosome breakpoints, revealed hot spots that could not be related to DNA concentration. It was one such hot spot in which the five identical X-Y translocations described by Merriam were found. There is not sufficient information to say that breaks in hot spots tend to interact preferentially with those in other hot spots, as would be expected if sequential affinities were responsible. The detailed analysis of induced rearrangements currently being pursued by Lefevre (1973) represents a powerful tool in probing the organization of *Drosophila* chromosomes. The fraction of breaks that are heterochromatic is similar to the fraction of the complement that is heterochromatic; however, Muller (1945) observed that an X chromosome lacking three-fourths of the proximal heterochromatin is not involved in X-autosome translocations appreciably less frequently than an X that has that same heterochromatic material present twice. This observation was confirmed in both Muller's (Wagoner, 1965) and my (Williams, 1965) laboratories (Table 2). The various chromosomes used have two kinds of heterochromatic segments, each of which may be either of two types: one contains approximately three times the amount of DNA of the other, and either size segment may be located either proximally near the centromere or distally near the tip of the chromosome. Williams performed a cytological analysis on the majority of his translocations and showed that the frequencies of breaks induced in large and small heterochromatic segments were 0.025 (10/395) and 0.024 (10/417) respectively, and the frequencies in proximal and distal segments were 0.015 (6/411) and 0.036 (15/411), versus 0.036 (15/411) in the euchromatin of the X chromosome. These observations are all consistent with Muller's hypothesis that the segment being manipulated in these experiments does not sustain rearrangement breakpoints and that the overall proportionality between breakpoints and quantity of heterochromatin is the result of an extra-

TABLE 2. The relation between the incidence of translocations involving the X chromosome and X chromosome length following irradiation of mature sperm with 4 kr of X-rays

X chromosome	Relative length*	Irradiated gametes tested	T(1;2)	T(1;3)	T(1;2;3)	T(2;3)	%T(X;A)	T(X;A) T(2;3)	Reference
$In(1)sc^{YL}sc^{8R}$	0.75	57	1	2	2	8	8.8	0.375	Williams (1965)
$In(1)sc^{4L}sc^{4R}$	1.00	156	7	10	5	9	14.1	1.89	
$In(1)sc^{S1L}sc^{8R}$	1.00	191	7	9	5	16	11.0	1.0	
$In(1)sc^{S1L}sc^{4R}$	1.25	134	5	8	1	9	10.4	1.44	
$In(1)sc^{4L}sc^{S1R}$	0.75	1170	43	81	30	150	13.2	0.827	Wagoner (1965)
$In(1)sc^{S1L}sc^{4R}$	1.25	1272	94	70	53	169	17.1	0.970	

*Relative length computed on the basis of Rudkin's (1965) estimate that the constitutive heterochromatin comprises one-third the X chromosome and Cooper's (1959) estimate that the region between the right breakpoints of $In(1)sc^4$ and either $In(1)sc^8$ or $In(1)sc^{S1}$ occupies three-fourths of the constitutive heterochromatin of the X.

ordinarily high breakability in the quarter of the heterochromatin not manipulated in these experiments. Thus, the heterochromatin sustains its share of the breaks, but their distribution within the heterochromatin is highly non-random.

Cooper and Krivshenko's unpublished cytological observation that heterochromatic breakpoints in free-X duplications are uniformly distributed within the proximal heterochromatin is incompatible with Muller's hypothesis. It seems to me more likely that these results are explicable in terms of the distribution of chromosomes within the male pronucleus than the distribution of X-ray induced lesions within the complement.

The final facet of chromosome organization that I wish to discuss is the nature of chromosome ends. Ends of chromosomes are stable in that they never rejoin with other chromosome ends or lesions, and they must therefore have a special structure. The creation of new chromosome ends is apparently possible under some but not all conditions. McClintock (1951) has shown that the fate of broken ends produced by the rupture of anaphase bridges in the gametophyte generation and the endosperm of corn is such as to produce bridges in the subsequent anaphase. It is as though the two broken strands of the double helix rejoin with one another in a hairpin configuration and at replication produce a dicentric chromatid whose centromeres are sisters which separate from one another at anaphase; this results in a chromatid type bridge-breakage-fusion cycle. Similar behavior has been observed (Peacock et al., 1974) in chromosome breaks induced in Chinese hamster ovary tissue culture cells by X-irradiation during G1 and can be inferred for X-ray induced chromosome breaks in mature sperm in *Drosophila* from the absence of demonstrable terminal deletions (e.g., Raffel, 1939). In the maize embryo on the other hand, products of anaphase bridge rupture form stable ends. Similar behavior of X-ray induced breaks in *Allium cepa* and *Vicia faba* root tip cells has been observed by Geard (personal communication). McClintock has contrived two products of anaphase rupture in the same cell; in gametophyte and endosperm these remain separate and generate two *chromatid* cycles, but in the zygote they invariably rejoin with one another to generate a *chromosome* bridge-breakage-fusion cycle in succeeding mitoses. The molecular nature and behavior of newly induced broken ends as well as the differences between these and normal chromosome ends pose fascinating problems for future consideration.

REFERENCES

- Abrahamson, S., M.A. Bender, A.D. Conger and S. Wolff. 1973. The uniformity of radiation-induced mutation rates among different species. *Nature* 245: 460.

- Bauer, H., M. Demerec and B.P. Kaufmann. 1937. X-ray induced chromosomal aberrations in *Drosophila melanogaster*. *Genetics* 23: 610.
- Brewen, G. and W.J. Peacock. 1968. Restricted rejoining of chromosomal subunits in aberration formation: A test for subunit dissimilarity. *Proc. Nat. Acad. Sci. USA* 62: 389.
- Brown, S.W. 1966. Heterochromatin. *Science* 151: 417.
- Champoux, J.J. and R. Dulbecco. 1972. An activity from mammalian cells that untwists superhelical DNA - a possible swivel for DNA replication. *Proc. Nat. Acad. Sci. USA* 69: 143.
- Cooper, K.W. 1959. Cytogenetic analysis of major heterochromatic elements (especially Xh and Y) in *Drosophila melanogaster* and the theory of 'heterochromatin'. *Chromosoma* 10: 535.
- Crick, F. 1971. General model for the chromosomes of higher organisms. *Nature* 234: 25.
- Cryer, D.R., C.D. Goldthwaite, S. Zinker, K.B. Lam, E. Storm, R. Hirschberg, J. Blamire, D.B. Finkelstein and J. Marmur. 1974. Studies on nuclear and mitochondrial DNA of *Saccharomyces cerevisiae*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 17.
- Gelbhart, W.M., M. McCarron, J. Pandey and A. Chovnick. 1974. Genetic limits of the xanthine dehydrogenase structural element within the rosy locus in *Drosophila melanogaster*. *Genetics*. In press.
- Hotta, Y. and H. Stern. 1974. DNA scission and repair during pachytene in *Lilium*. *Chromosoma*. In press.
- Ives, P.T. 1972. Residual lethality in induced rearrangements and the chromosomal distribution of male crossovers. *Dros. Info. Serv.* 49: 94.
- Kaufmann, B.P. 1946. Organization of the chromosome. I. Break distribution and chromosome recombination in *Drosophila melanogaster*. *J. Exp. Zool.* 102: 293.
- Kavenoff, R. and B.H. Zimm. 1973. Chromosome sized DNA molecules from *Drosophila*. *Chromosoma* 41: 1.
- Laird, C.D. 1973. DNA of *Drosophila* chromosomes. *Adv. Genetics* 7: 177.
- Lefevre, G. 1971. Salivary chromosome bands and the frequency of crossing over in *Drosophila melanogaster*. *Genetics* 67: 497.
- . 1974. The one band-one gene hypothesis: Evidence from a cytogenetic analysis of mutant and nonmutant rearrangement breakpoints in *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 591.
- Lifschytz, E. and D.L. Lindsley. 1972. The role of X-chromosome inactivation during spermatogenesis. *Proc. Nat. Acad. Sci. USA* 69: 182.

- Lodish, H.F., R.A. Firtel and A. Jacobson. 1974. Transcription and structure of the genome of the cellular slime mold *Dictyostelium discoideum*. *Cold Spring Harbor Symp. Quant. Biol.* 38: 899.
- McClintock, B. 1951. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* 16: 13.
- Muller, H.J. 1945. The non-equivalence of the blocks and the salivary 'heterochromatin'. *Genetics* 30: 15.
- Olins, A.L. and D.E. Olins. 1974. Spheroid chromatin units (v Bodies). *Science* 183: 330.
- Peacock, W.J., S. Wolff and D.L. Lindsley. 1973. Continuity of chromosome subunits. In *Chromosomes Today* (J. Wahrman and K.R. Lewis, ed.) Vol 4, p 85. John Wiley, New York.
- Petes, T.D. and W.L. Fangman. 1972. Sedimentation properties of yeast chromosomal DNA. *Proc. Nat. Acad. Sci. USA* 69: 1188.
- Raffel, D. 1939. A genetic analysis of apparent losses of the distal end of the scute-8 chromosome. *Genetics* 24: 107.
- Rudkin, G.T. 1965. The structure and function of heterochromatin. In *Genetics Today, Proc. XI Int. Congr. Genet.* (S.J. Geerts, ed.) Vol 2, p 359. Pergamon Press, Oxford.
- Sandler, L. and F. Hecht. 1973. Genetic effects of aneuploidy. *Amer. J. Human Genet.* 25: 332.
- Taylor, J.H. 1958. Sister chromatid exchanges in tritium-labeled chromosomes. *Genetics* 43: 515.
- Taylor, J.H., P.S. Woods and W.L. Hughes. 1957. The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium labeled thymidine. *Proc. Nat. Acad. Sci. USA* 43: 122.
- Wagoner, D.E. 1965. The frequency of effective breakage of heterochromatin in *Drosophila melanogaster*. *Thesis*, University of Indiana.
- Wang, J.C. 1971. Interaction between DNA and an *Escherichia coli* protein ω . *J. Mol. Biol.* 55: 523.
- Williams, B.J. 1965. The effects of the amount and distribution of heterochromatin on the frequency of X-ray induced X-autosomal translocations in *Drosophila melanogaster*. *Thesis*, University of Tennessee.
- Woodcock, C.L.F. 1973. Ultrastructure of inactive chromatin. *J. Cell. Biol.* 59: 368a.

Subject Index

- Acidic proteins
 chromatin, 255-65, 273-9, 477
- Allozymes, 381-91
 polymorphism, 383
- Aplysia californica*, 74
- Bos domesticus*, 74
- Calliphora stygia*, 149-54
- Cell
 cycle, 369-71, 395-6, 481
 fusion, 367-78
 haploid, 451-3
 human fibroblast, 459-69
 hybridization, 367-78
 plant culture, 451-5
 protoplasts, 448, 453
- Centromere, 51-3, 56-8, 314,
 317, 475, 482
- Chiasmata, 321, 323, 487
- Chloroplasts, 236
- Chromatid
 polarity, 473-4
 sister, 473, 483
 unineme, 474
- Chromatin, 245-52, 255-65,
 269-80, 314-23, 477
 acidic proteins, 255-68,
 273-9, 477
 centromeric, 315
 chick, 255-65
 fractionation, 265
 histones, 257-65, 272-9
 model, 247, 477
- Chromocentre, 22, 27, 51-3,
 56-8, 483
- Chromomere, 1-30, 169-82
 cistron, 2, 4
 DNA content, 170, 181
 duplication, 26
 size, 170, 181
- Chromosome
 axes, 275
- Chromosome (cont'd)
 band, 1, 15-23, 51-3, 170-82
 bineme, 123, 129-32
 breakage, 362-4, 483, 490,
 492
 centromere, 51-3, 56-8, 314,
 317, 475, 482,
 484-5, 489-90
- Cercatetus*, 390
- chiasmata, 321, 323
- chromocentre, 22, 27, 51-3,
 56-8, 483
- condensation, 477, 483-4
- crossing over, 283-99, 301-
 10, 317-23
- dicentric, 492
- DNA content, 130-1, 327
- dosage compensation, 382,
 390-1
- Drosophila*, 1-30, 35-44, 47-
 58, 87-107, 169-
 82
- ends, 314-5, 492
- euchromatin, 396-407, 475
- evolution, 395-407
- fragment, 483-4
- function, 477
- grasshopper, 395-407
- hamster, 367-78, 482
- heterochromatin, 27, 35, 44,
 51-3, 56-8,
 396-407,
 475, 484-5,
 489, 490-2
- heteropycnosis, 400
- interband, 1, 15-23
- kangaroo, 381-91
- karyotype, 368-78, 395, 404-5
- knobs, 316-7
- late-labelling, 399-407, 485
- lampbrush, 277, 477
- Lilium*, 283-99
- length, 389-90
- mammalian, 355-65, 381-91
- marsupial, 381-91

Chromosome (cont'd)

- Macropus*, 381-91
- model, 2-4, 87-107, 123-32, 144, 192, 247, 364-5, 473-4, 477, 481-4
- mouse, 367-78, 482
- movement, 477, 485
- megameric, 397-9
- Microtus*, 390
- oraba*, 395-407
- nonhomologous pairing, 321
- Neurospora*, 301-10
- newt, 270, 274-80
- nucleolus organiser, 58, 315, 317, 320
- number, 395, 403
- organisation, 61-85, 473-92
 - periodicity, 487-8
- pairing, 44, 301, 317
- polytene, 1, 21, 51-3, 149, 169-82, 279, 482-4
- polyneme, 124, 132
- Phascolomis*, 389
- proteins, 246-52, 255-65, 269-80, 477
- rearrangements, 176-8
- replication, 123-32, 144, 327-34, 355-65, 369-78, 395-407, 477, 481-3
- reunion, 483
- ring, 473
- sea urchin, 269-74
- synapsis, 283-99, 313-23, 403, 485-6
- synaptonemal complex, 285, 313-23, 485, 487
- translocation, 388-90, 399, 478-80, 490-1
- transcription, 477-8
- unineme, 123-32
- X-inactivation, 381-91, 485
- yeast, 327-42, 476
- Colymbetes fuscus*, 138
- Conversion, gene, 301, 486

Conversion (cont'd)

- polarity of, 306-8
- Crossing over, 283-99, 301-10, 317-23
- Diplochromosomes, 126, 132
- DNA
 - Acetabularia*, 143
 - amplification, 139-45, 156-7, 163-4, 482
 - Arabidopsis*, 449-50
 - arrangement, 474, 487-9
 - Calliphora*, 149-54
 - chain termination, 356-65
 - Chlamydomonas*, 143
 - Colymbetes*, 138
 - chromosome location, 15-25, 27, 36, 44, 51-3, 57, 97-101, 144, 288
 - chromosomal molecule, 327, 335-41, 476
 - circular, 4, 7, 15, 87, 104-5, 111-22, 138, 164, 345-53, 410, 426
 - cleavage, 434-44
 - deletion, 155-64, 412-24, 431-43
 - density, 9, 14-15, 36, 47, 90, 96, 163, 285, 449
 - developmental time, 150-2
 - differential replication, 150-2
 - DNA hybridization, 294, 474-6
 - Drosophila*, 1-30, 35-44, 47-58, 75, 87-107, 474-6, 479
 - duplication, 418
 - Dytiscus*, 137-8
 - E. coli*, 90-104, 200, 451-5
 - evolution, 36, 44
 - folding, 245
 - fragments, 476
 - function, 44, 58
 - genome size, 327

DNA (cont'd)

g-regions, 87-107
 heteroduplex, 9-15, 27, 414,
 426, 431-43
 heterogeneity, 39, 43-4
 histone, 249-52, 477
 homoduplex, 438-40
 homogeneity, 43-4
 hybrid, 4-30
 insertion, 425-7, 431-43,
 interspersed repeats, 3, 63-
 73
 λ , 91-5, 451-5, 459-69
 length, 410, 475
 mammalian 355-65
Micrococcus, 449
 minicircular, 111-22
 mitochondrial, 112-22, 155-69
 mouse, 74, 200-1, 355-65
 mutation mapping, 431-43
 non-repeated sequence, 15-21,
 23-30,
 62-85,
 193
 Okazaki fragments, 348-53,
 481
 oDNA, 112-15
 P-DNA (pachytene), 284, 295-9
Petunia, 448-9
 ϕ 80, 451-5
Physarum, 142-3
 plasmid, 4-30
 polytene chromosome, 149
Polyoma, 345-53, 481
 pyrimidine sequence, 365
 pyrimidine tract, 200-2,
 reassociation kinetics, 15-24,
 284,
 287-99
 regulation, 156-7
 repair replication, 298, 357
 487
 repeating units, 245-52
 repeated sequence, 3, 15-21,
 23-30, 35-44, 47-58,
 62-85, 87-107, 193,
 285-99, 310, 474-5

DNA (cont'd)

replication, 136-9, 142, 149-
 54, 327-34, 345-
 53, 481-2, 369-
 71, 395, 407
 replicative intermediate, 345-
 53
 ribosomal, 28, 115, 135-45,
 149-54, 156-67, 482
 amplification, 139-45,
 156-7, 163-4,
 482
 chromosomal location, 144
 copies, number, 141
 extrachromosomal, 135-45,
 482
 integration, 141, 144
 oocytes, 135-7
 replication, 138-9, 142
 restriction endonuclease, 4-
 30, 88-107, 411-
 27, 433-44, 476
 ρ DNA, 112
 RNA hybridization, 77-85, 142,
 152, 59-
 62, 193,
 206-12, 270
 474-6
Saccharomyces, 112-22, 327-42,
 476
 satellite, 474, 485, 487-8
 segment length, 47-51, 56
 sea urchin, 61-85, 193, 475
 sequence, 35-44, 70-1
 arrangement, 53-6, 62-85
 diversity, 474
 homology, 294
 length, 62-3, 67, 87, 90-107
 single copy, 62-85, 288
 single stranded, 418
 single strand nuclease (S1)
 62-71, 212, 431-43
 spacer, 28-9, 140
Stylonychia, 142-3
 supercoiled, 349-53, 412, 414
 SV40 410-27, 431-43
 synthesis, 369-71

DNA (cont'd)

- tandem array, 474
- tandem repeats, 3, 18, 25-6, 87-107
- Tetrahymena*, 139-45, 482
- under-replication, 482-3
- Xenopus*, 61-85, 135-8, 140, 143, 193
- yeast, 112-22, 327-42, 476
- Z-DNA (zygotene), 284-95
- Drosophila melanogaster*, 1-30, 35-44, 47-58, 75, 87-107, 169-82, 474-6, 479
- E. coli* hybrid DNA, 4-30
- mutants
 - zeste*, 171, 175, 178-9
 - white*, 171, 173-80
 - roughest*, 178, 179
 - verticals*, 178, 179
- Dytiscus marginalis*, 137-8
- Endoreduplication, 125-32

Enzyme

- allozymes, 381-91
- bacterial in human cells, 461-3
- galactosemia, 459-69
- markers, 301-10, 359-61, 381-91
- nuclease (S1), 62-71, 212, 431-44
- restriction, 4-30, 88-107, 411-27, 433-44, 476
- Evolution, 36-44, 395-407
- Escherichia coli*, 4-30, 90-104, 200, 451-5, 474-6, 479
- D. melanogaster*, hybrid DNA, 4-30
- plasmid, 4-30
- Feather keratin, 205-13
- Galactosemia, 459-69

Gene

- bacterial, function in plant cells, 447-55
- bacterial, function in human cells, 461-3
- complementation, 170-82
- conversion, 301, 306-8, 486
- duplications, 173-8
- enzyme markers, 301-10, 359-61, 381-91
- expression, 382, 448-55, 459-69
- feather keratin, 205-13
- multiple homologous, 206
- recombination, 171
- regulation, 170-82, 185-93
- reiteration frequency, 212
- sex linked, 381-91
- structural, 73-85
- suppressor, 468-9
- transcription, 185-93, 257, 263, 265, 477-81
- transgenesis, 447-55, 459-69
- translation modulation, 187
- transposition, 178-80
- Globin, 257, 265
- Grasshopper, 395-407
- Heterochromatin, 27, 35-45, 47-59, 396-407, 475, 484-5, 489, 490-2
- Histones, 246-52, 257-65, 272-9, 477
 - polymer, 246-52
 - repeating arrangement, 246-51
 - tetramers, 246, 250
- Hybrids, 367-78
 - cell, 367-78
 - mouse x hamster, 367-78, 482
- Ilyanassa obsoleta*, 75
- Kangaroo, 381-91
- Keratin, 205-13
 - embryo chick feather, 205-13
 - synthesis, 205-6
- Kilobase (kb), 1, 473
- λ , 91-5, 451-5, 459-69

- Lilium*, 283-99
Lycopersicon esculentum, 451-4
 Meiosis, 313-23
 Mitochondria, 112-22, 155-69
Moraba cultrata, 400-3
Moraba viatica, 398-9
Moraba virgo, 403-7
Mus musculus, 74, 200-1, 218-30,
 355-65, 367-78,
 482
 Mutagenesis, 329-34, 481
 frequency, 362-4, 478
 lethal, 478-80
 replication dependent, 330,
 355-65, 481
Neurospora crassa, 301-10,
 486-7
 Newt, crested, 270, 274-80
 Nuclear envelope, 3, 4-5, 314-
 5, 321
 Nucleolus, 22, 28-9, 58, 144-5,
 152-3, 275-9, 315-
 20, 390
 organisation, 29
 organizer region, 315-20
Oncopeltus fasciatus, 75
 Okazaki fragment, 348-53, 481
 Parthenogenesis, 403-7
 Plasmid, 4-30
 Polyoma virus, 345-53, 481
 Polyploidy, 125-32
 Polysomes, 236-7, 257
 Protein synthesis, 149-53,
 238-9
Rattus norvegicus, 74
 Recombination, 283-99, 301-10,
 486
 allelic, 303-10
 chiasmata, 321-3
 chiasma frequency, 487
 control genes (*con*), 302-10,
 486
 frequency, 302-10, 486
 gene, 171
 mitochondrial DNA, 116-20
 pairing, 317
 Recombination (cont'd)
 recognition gene (*cog*), 309-10
 recombination gene (*rec*),
 302-10, 486-7
 synapsis, 283-99, 314, 485-6
 Regulatory genes, 170-82, 188-
 92
 Replication, 327, 345-53, 477,
 481-3
 chromosome, 123-32, 144, 327-
 34, 355-65, 369-
 78, 395-407, 477,
 481-3
 control, 367-78, 481
 DNA, 136-9, 142, 149-54,
 327-41, 345-53, 481-2,
 369-71, 395, 407
 initiation, 327-34, 345,
 481-2, 489
 intermediates, 345-53
 late, 399-407, 484-5
 mapping, 357-62
 Okazaki fragment, 348-53,
 481
 repair, 298, 357, 487
 sequence, 373-5, 397, 481
 termination, 345, 356-65
 Restriction endonuclease, 4-30,
 88-107, 411-27,
 433-44, 476
 Eco RI, 4-30, 88-107, 411-27,
 433-44
 Eco RII, 412
 Hae, 88-107
 Hind, 88-107, 411-27
 Hpa I, II, 411-27, 433-44
 RNA
 base sequence, 234
Dyctiostelium, 476
 DNA hybridization, 77-85, 152,
 159-62,
 193, 206-
 12, 270
 double stranded (dsRNA), 217-30

RNA (cont'd)

- HnRNA, 77-9, 186-93, 202, 222-9, 234, 475, 478-81
- heterogeneity, 235
- mRNA, 73-9, 185-93, 199, 201-2, 206-12, 218-30, 233-9, 257, 475-6
- mammalian, 185-93
- mouse, 218-30
- nRNA, 219, 222-4, 234
- plant, 233-9
- polysomal, 218, 223, 225
- post-transcriptional modification, 186, 199, 201-2
- rRNA, 28, 185, 219, 225
- rapidly labelled, 234-5
- rat, 218-30
- regulation, 187, 263
- repetitive sequences, 229
- RNA hybridization, 142, 219, 222-30
- size, 475
- synthesis, 186-93, 235-6
- tRNA, 185, 219, 224-5
- Ribosomes, 236
- Saccharomyces cerevisiae*, 111-22, 156-64, 327-42, 476
- Sea urchin, 61-85, 193, 269-74, 475
- Simian virus 40*, 409-27, 431-44, 454
- Strongylocentrotus purpuratus*, 61-85, 193, 269-74, 475
- Structural genes
 - DNA sequence, 73-7
 - size, 77-9
 - number, 79-81
 - proximity to repeated sequence, 81-5
- Synaptonemal complex, 285, 313-23, 485, 487

Tetrahymena pyriformis, 139-45, 482

Tomato, 451-4

Transgenesis

- DNA mediated, 447-55
- phage mediated, 451-5, 459-69
- Transcription, 185-93, 270, 477-81
 - avian erythroid cells, 257
 - control, 478-81
 - regulation, 263, 265
 - modification, 186, 199, 201-2
- Triturus cristatus*, 270, 274-80
- Urechis caupo*, 75

Virus

- defective, 412-27
- DNA, 410-27
- genome mapping, 409-27, 431-44
- λ , 91-5, 451-5, 459-69
- phage, 447, 451
- ϕ 80, 451-5
- Polyoma, 345-53, 481
- segments, 412-27
- specialised transducing, 447, 451-5, 459-69
- SV40, 410-27, 431-44, 454
- T3, 453, 468-9
- Xenopus laevis*, 26, 28-9, 61-85, 135-38, 140, 143, 193
- Yeast, 111-22, 327-42, 476
 - petite, 111-22, 156-64

Cover photograph

Shows localization of a highly repeated
DNA sequence in *Drosophila melanogaster*

Australian National University Press Canberra

ISBN 0 7081 0104 6

