

DISS. ETH NO. 17313

STRUCTURAL AND BIOCHEMICAL STUDIES ON HUMAN
TAF2

A dissertation submitted to
ETH Zurich

for the degree of
Doctor of Natural Sciences

presented by
MARTINA MIJUSKOVIC
Dipl. Eng. Biol., Univ. of Zagreb, Croatia
born March 26, 1979
citizen of Croatia

accepted on the recommendation of
Prof. Timothy J. Richmond, examiner
Prof. Rudolf Glockshuber, co-examiner

2007

ACKNOWLEDGEMENTS

I would like to thank Prof. Timothy J. Richmond for giving me the opportunity to work in his laboratory and Dr. Imre Berger for guiding me through first years of my graduate work.

I am very grateful to Prof. Rudi Glockshuber for critical reading of this thesis as a co-examiner and all his helpful comments.

I am thankful to experts from the Functional Genomic Center Zurich for providing the protein analysis service crucial for this work. Mike Scott deserves a very special thanks for introducing me to the Biacore technology and assisting with all measurements and data analysis.

I would like to thank Beat Blattmann from the University of Zurich for providing an excellent protein crystallization service.

I would like to thank all current and previous lab members for providing a pleasant work environment. In particular, I am very grateful to Dr. Daniel Fitzgerald for his critical comments and readiness to help at most difficult times. I am also especially grateful to Dr. Benedetta Dorigo for her support throughout these years. Dr. David Sargent deserves a very special thanks for his assistance with many technical aspects of this work. I am also thankful to Dr. Kazuhiro Yamada for helpful discussions and his kind support. Dr. Sabrina Kammerer deserves my gratitude for her support and helping me with the French translation of the Summary. Finally, I am most thankful to Sylwia Duda for not only being a lab-mate but also a wonderful friend.

My deepest gratitude goes to my friends Branimir, Dunja, Milica and Sanja for their patience and unconditional support through all these years. Also, I am grateful to all participants of the Summer School of Science in Visnjan, Croatia, for preserving my enthusiasm and motivation to do research.

At the end, I would like to thank Ivana Frajtic-Mijuskovic, for teaching me that I can do anything I want to do and for being the most wonderful and supportive mother any daughter could wish for.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	I
SUMMARY	V
RÉSUMÉ	VI
ABBREVIATIONS	VIII
1 INTRODUCTION	1
1.1 Transcription: first stage of gene expression	1
1.1.1 Central dogma of molecular biology	1
1.1.2 The role of RNA in protein synthesis	2
1.2 RNA polymerases	3
1.3 RNA polymerase II core promoter	6
1.3.1 Initiation of transcription in Eukaryotes	6
1.3.2 Core promoter elements.....	6
1.3.2.1 The TATA box.....	8
1.3.2.2 The initiator (Inr)	9
1.3.2.3 The downstream core promoter element (DPE)	11
1.3.2.4 The TFIIB recognition element (BRE).....	11
1.3.2.5 CpG islands.....	12
1.3.3 Transcriptional regulation via core promoter elements	12
1.4 General transcription factor TFIID	14
1.4.1 TBP-associated factors (TAFs).....	14
1.4.2 TFIID architecture	16
1.4.3 Other TAF-containing complexes.....	18
1.5 Outline of the thesis	19
2 MATERIALS AND METHODS.....	20
2.1 Materials, chemicals and equipment.....	20
2.1.1 General chemicals	20
2.1.2 Buffers and solutions	21
2.1.3 Enzymes.....	23
2.1.4 Bacterial strains and media	24
2.1.5 Insect cell culture material and equipment	26
2.1.6 Cloning and expression vectors	26
2.1.7 Oligonucleotides	27
2.1.8 DNA and protein markers.....	31
2.1.9 Commercial kits	31
2.1.10 Equipment.....	32
2.2 Recombinant DNA methods.....	35
2.2.1 Determination of DNA concentration.....	35
2.2.2 DNA polyacrylamide gel electrophoresis (DNA PAGE)	35
2.2.3 Agarose gel electrophoresis	35
2.2.4 Small scale alkaline lysis plasmid preparation (“miniprep”).....	35
2.2.5 Medium scale alkaline lysis plasmid preparation (“midiprep”)	36
2.2.6 Ethanol precipitation of DNA fragments.....	36

2.2.7	Spin column purification.....	37
2.2.8	Agarose gel purification.....	37
2.2.9	Preparation of electrocompetent cells.....	37
2.2.10	Transformation of bacteria by heat shock.....	38
2.2.11	Transformation of bacteria by electroporation	38
2.2.12	Restriction endonuclease digestion.....	38
2.2.13	DNA dephosphorylation	39
2.2.14	Ligation.....	39
2.2.15	DNA sequencing.....	39
2.2.16	Restriction fragment subcloning.....	40
2.2.17	PCR subcloning	40
2.2.18	Design of PCR primers	41
2.2.19	PCR subcloning using the TOPO TA cloning strategy	42
2.2.20	Gene SOEing method	42
2.2.21	Quick change mutagenesis (adapted from S. Hahn lab).....	42
2.2.22	Annealing of the oligonucleotides	43
2.2.23	DNA end labeling with ³² P	43
2.3	Bacterial protein expression methods	45
2.3.1	Small scale expression	45
2.3.2	Large scale expression	45
2.3.3	Solubility test	46
2.4	Insect cell (baculo) expression methods	47
2.4.1	Initiating a cell culture from a frozen stock.....	48
2.4.2	Maintaining a cell culture in suspension.....	48
2.4.3	Production of recombinant bacmid DNA	48
2.4.4	Transfection of Sf21 cells.....	50
2.4.5	Virus amplification and maintenance	50
2.4.6	Plaque assay for virus titer determination.....	50
2.4.7	Expression tests.....	51
2.4.8	Small and medium scale expression in shaker flasks	51
2.4.9	Large scale expression in the bioreactor.....	52
2.4.10	Preparation of nuclear and cytosolic insect cell extracts for SDS PAGE analysis.....	52
2.5	Protein purification methods.....	54
2.5.1	Preparation of protein extract (cell lysis).....	54
2.5.1.1	Bacterial cell lysis.....	54
2.5.1.2	Preparation of insect cell cytoplasmic and nuclear fraction	54
2.5.2	Batch protein purification	55
2.5.3	Fast protein liquid chromatography (FPLC).....	55
2.5.4	Production of TEV protease.....	55
2.5.5	Production of core human TBP.....	56
2.5.6	Production of human TAF8/10 complex	58
2.6	Protein analysis methods.....	60
2.6.1	Determination of protein concentrations	60
2.6.2	Denaturing polyacrylamide gel electrophoresis (SDS-PAGE).....	60
2.6.3	Western blotting.....	61
2.6.4	N-terminal sequencing	62

2.6.5	Mass spectrometry	62
2.6.6	Limited proteolysis	62
2.6.7	Electrophoretic mobility shift assay.....	63
2.6.8	Analytical ultracentrifugation (sedimentation velocity experiments).....	63
2.7	Circular dichroism (CD) spectroscopy	65
2.7.1	Thermal stability measurements	66
2.7.2	Measurements of secondary structure content.....	67
2.8	Surface plasmon resonance (BiaCore) methods	69
2.8.1	The principle of the surface plasmon resonance (SPR).....	69
2.8.2	BiaCore sensor chips.....	71
2.8.3	Experimental setup.....	71
2.8.4	Protein immobilization by amine coupling.....	73
2.8.5	Protein immobilization by thiol coupling	73
2.8.6	Biotinylated DNA immobilization.....	73
2.8.7	Data analysis	74
2.9	Crystallization screens and methods.....	75
2.9.1	General remarks	75
2.9.2	Vapor diffusion crystallization	75
2.9.3	Crystallization using the Cartesian robot.....	75
2.9.4	Crystallization screens	76
2.10	Bioinformatics tools, databases and software	77
3	RECOMBINANT PRODUCTION AND PURIFICATION OF FULL-LENGTH HUMAN TAF2	78
3.1	Introduction.....	78
3.2	Expression of hTAF2 in Sf21 insect cells.....	78
3.2.1	Virus production and expression tests	78
3.2.2	Large scale expression	81
3.3	Purification of hTAF2.....	83
3.3.1	Cytoplasmic localization of hTAF2.....	83
3.3.2	Initial purification trials	83
3.3.3	Q-sepharose anion-exchange chromatography	87
3.3.4	TALON™ affinity chromatography	87
3.3.5	Resource S cation-exchange chromatography	90
3.3.6	Size-exclusion chromatography.....	92
3.3.7	Analysis of hTAF2 by analytical ultracentrifugation	94
3.3.8	hTAF2 purification overview	95
3.4	Discussion.....	98
4	DEFINITION AND STRUCTURAL CHARACTERIZATION OF hTAF2 DOMAINS .	100
4.1	Introduction.....	100
4.2	hTAF2 domain definition	101
4.2.1	Analysis of hTAF2 primary structure by bioinformatics.....	101
4.2.2	Analysis of full-length hTAF2 by CD spectroscopy	106
4.2.2.1	Melting temperature determination.....	106
4.2.2.2	Analysis of secondary structure content	107
4.2.3	Limited proteolysis of full-length hTAF2.....	108
4.2.4	Model of hTAF2 domain structure	111

4.3	Production and characterization of the hTAF2 core domain	113
4.3.1	Design, production and the analysis of hTAF2 “TEV” constructs.....	113
4.3.1.1	Cloning of hTAF2 “TEV” constructs	114
4.3.1.2	Expression and purification of hTAF2 TEV construct A	115
4.3.1.3	Expression and purification of hTAF2 TEV construct B	118
4.3.1.4	Expression and purification of hTAF2 TEV constructs AB, D, AD and D2..	121
4.3.2	Analysis of the hTAF2 TEV construct A	124
4.3.2.1	Limited proteolysis experiments.....	124
4.3.2.2	Analysis of endogenous proteolytic degradation fragments.....	129
4.3.2.3	Gel filtration experiments with proteolytic mixtures.....	131
4.3.2.4	Ion-exchange chromatography with tryptic fragments	133
4.4	Production and characterization of hTAF2 tail domain.....	136
4.4.1	Cloning, bacterial expression and purification of hTAF2 tail domain	136
4.4.2	Analysis of endogenous degradation fragments	140
4.4.3	Structural analysis by CD spectroscopy	143
4.5	Discussion	144
5	FUNCTIONAL CHARACTERIZATION OF hTAF2.....	147
5.1	Introduction.....	147
5.2	Production of hTAF1 and attempts to analyze the hTAF1-hTAF2 interaction.....	149
5.2.1	Recombinant production and purification of full-length hTAF1.....	149
5.2.1.1	Analysis of hTAF1 primary structure by bioinformatics.....	149
5.2.1.2	Cloning of the hTAF1 expression construct	154
5.2.1.3	Expression of hTAF1 in Sf21 insect cells.....	156
5.2.1.4	Purification of hTAF1	157
5.2.2	Production of truncated hTAF1 constructs and their co-expression with hTAF2	163
5.2.2.1	Cloning of hTAF1 constructs.....	165
5.2.2.2	Expression of hTAF1 constructs in Sf21 insect cells	166
5.2.2.3	Purification of hTAF1 C-terminal domain (construct 2) and hTAF2 interaction	166
	analysis	166
5.2.2.4	Co-expression of hTAF1 constructs 1 and 3 with hTAF2.....	168
5.2.3	Co-expression of hTAF1, hTAF2 and hTBPc in Sf21 insect cells.....	169
5.3	Interaction of hTAF2 with hTAF8/10.....	172
5.3.1	Interaction of full-length hTAF2 with hTAF8/10 complex	172
5.3.2	hTAF8/10 binding activity of hTAF2 domains	176
5.4	hTAF2 DNA binding activity	180
5.4.1	Analysis of full-length hTAF2 DNA binding activity	180
5.4.2	DNA binding activity of hTAF2 domains	183
5.5	Discussion and perspectives	187
6	REFERENCES	190
	APPENDIX I: DNA AND PROTEIN SEQUENCES	198
	APPENDIX II: EXPRESSION OF TAF2 CONSTRUCTS IN <i>E. COLI</i>	210
	APPENDIX III: TAF2 CRYSTALLIZATION ATTEMPTS – SUMMARY AND	
	PERSPECTIVES	213
	CURRICULUM VITAE.....	214

SUMMARY

Eukaryotes have thousands of genes whose expression must be tightly regulated to ensure the organism's survival, growth and development. Synthesis of messenger RNA, a process known as transcription by RNA polymerase II, is a first step in the expression of protein-coding genes. RNA polymerase II core promoter is the most important element in the control of transcription initiation. It is the ultimate target of action of all factors that are involved in the regulation of class II gene transcription. For RNA polymerase II to accurately and efficiently initiate transcription from the core promoter, it requires a set of "general transcription factors". These include the transcription factors (TF) IIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH. TFIID is a multisubunit complex composed of the TATA-box binding protein (TBP) and up to 14 TBP-associated factors (TAFs). It has a variety of functions, including core promoter recognition and enzymatic activities that are thought to assist in transcription activation.

The focus of this work is human (h) TAF2, the second largest subunit of human TFIID which is implicated in the recognition of initiator core promoter element together with hTAF1. The purification protocol of recombinantly produced hTAF2 is developed that provides enough material for crystallization experiments and biochemical characterization of hTAF2. By combination of limited proteolysis and gel filtration studies, hTAF2 is shown to consist of two structural domains. One is a large globular domain of compact structure and 116 kDa in size. This domain represents the structural core of hTAF2. The other domain, termed "TAF2 tail", is structurally disordered and 22 kDa in size.

Finally, functional analysis of hTAF2 revealed that it has non-specific DNA binding activity, localized to the lysine-rich tail domain. This activity might add to the recognition of core promoter by TFIID, but can not be responsible for the specific sequence recognition itself. *In vitro* binding experiments with human TAF8/TAF10 complex showed that hTAF2 directly interacts with this complex. This result reveals important new information about the architecture of TFIID.

RÉSUMÉ

Les organismes eucaryotes possèdent des milliers de gènes dont l'expression doit être finement régulée pour assurer leur survie, croissance et développement. La synthèse de l'ARN messenger, connue sous le nom de transcription dépendante de l'ARN polymérase de type II (ARN pol II), constitue une première étape dans le processus d'expression des gènes codant pour des protéines. La séquence cœur des promoteurs dépendants de l'ARN pol II est l'élément le plus important du contrôle de l'initiation de la transcription. Il est la cible ultime de l'action de tous les facteurs qui sont impliqués dans la régulation de la transcription des gènes de classe II. L'initiation de la transcription par l'ARN pol II, à partir de la séquence cœur du promoteur, requiert précision et efficacité ce qui nécessite le recours à une série de facteurs de transcription dits « généraux ». On compte les facteurs de transcription (TF, Transcription Factors) TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, et TFIIH. TFIID est un complexe multimérique composé de la protéine de liaison à la boîte TATA (TBP, TATA box Binding Protein), et jusqu'à 14 facteurs associés à la TBP (TAF, TBP Associated Factors). Il possède une variété de fonctions dont la reconnaissance du promoteur et des activités enzymatiques dont on pense qu'elles favorisent l'activation de la transcription.

Mon travail de thèse s'est focalisé sur TAF2 humain (hTAF2), la deuxième plus grande sous-unité de TFIID. hTAF2 est impliqué, avec hTAF1, dans la reconnaissance de l'élément initiateur de la transcription dans la séquence cœur du promoteur. Le protocole développé pour la purification de hTAF2, produit de manière recombinante, permet l'obtention de matériel en quantité suffisante pour la réalisation d'expérience de cristallisation et la caractérisation biochimique de hTAF2. De plus, la combinaison d'études par protéolyse ménagée et de filtration sur gel a montré que hTAF2 est constitué de deux domaines structuraux. L'un, le cœur structural de hTAF2, est un large domaine globulaire de structure compacte et de taille 116kDa. L'autre domaine, de taille 22 kDa, appelé « domaine queue de TAF2 » est structurellement désordonné.

Enfin, les études fonctionnelles réalisées ont révélé que hTAF2 possède une activité aspécifique de liaison à l'ADN, localisée dans le « domaine queue » riche en lysine. Cette activité pourrait favoriser la liaison de TFIID à la séquence cœur du promoteur, mais ne peut pas être responsable en elle-même de la reconnaissance spécifique du promoteur. Par ailleurs, les

tests de liaison réalisés *in vitro* ont montré que hTAF2 interagit directement avec le complexe TAF8/TAF10. Ce résultat révèle d'importantes informations novatrices concernant l'architecture de TFIID.

ABBREVIATIONS

A	absorbance
aa	amino acid
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumine
BLAST	basic local alignment search tool
cfu	colony forming unit
Da	Dalton (approx. g/mol)
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ds	doublestranded
EMSA	electrophoretic mobility shift assay
EtOH	ethanol
FPLC	fast protein liquid chromatography
GTF	general transcription factor
HPLC	high performance liquid chromatography
hTAF2	human TATA-box binding protein associated factor 2
kDa	kilodalton
MALDI-TOF	matrix-assisted laser desorption/ionisation – time of flight
MCS	multiple cloning site
MNase	micrococcal nuclease
MS	mass spectrometry
Mw	molecular weight
MWCO	molecular weight cut off
NCBI	National Center for Biotechnology Information
NLS	nuclear localization signal
NTP	nucleoside triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pfu	plaque forming unit
polh	polyhedrin promoter
RNA	ribonucleic acid
RNase	ribonuclease
rpm	rounds per minute
RT	room temperature (20-25° C)
SDS	sodium dodecyl sulphate
ss	single stranded
TAF	TATA-box binding protein associated factor
TBP	TATA-box binding protein

TFIID transcription factor II D
UV ultraviolet

Other abbreviations are listed in sections 2.1.1. (chemicals) and 2.1.2. (buffers and solutions).

1

INTRODUCTION*1.1 Transcription: first stage of gene expression*

1.1.1 Central dogma of molecular biology

DNA has long been known as the carrier of genetic information in all living organisms, directing the synthesis of proteins. However, the involvement of RNA in this process became evident only in the late 1930s, when the discovery was made of DNA being confined to the cell nucleus while the RNA is mostly cytosolic. In the 1950s, when radioactively labeled amino acids became available, cytoplasmic protein-RNA particles (ribosomes) were shown to be the sites of protein synthesis and the idea emerged that protein synthesis is not immediately directed by the DNA. In 1958, these relationships were summarized by Francis Crick in a flow scheme known as the central dogma of molecular biology: “*DNA directs its own replication and transcription to RNA which, in turn, directs its translation to proteins*” (Figure 1.1).

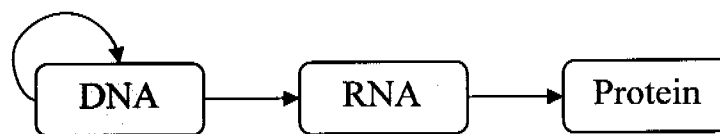


Figure 1.1. The central dogma of molecular biology.

The central dogma has been revised after the discovery of retroviruses, whose RNA genome can direct the synthesis of DNA through a special RNA-directed DNA polymerase (reverse transcriptase), as well as some RNA viruses with RNA-directed RNA-polymerases. However, information transfer in other directions never occurs, making proteins the only recipients of genetic information (Crick, 1970).

1.1.2 The role of RNA in protein synthesis

Observations made studying the *lac* operon in *E. coli* led to the conclusion that proteins are synthesized in a two step process: first, the structural genes on DNA are transcribed onto complementary strands of messenger RNA (mRNA) and second, mRNA is transiently associated with ribosomes directing them in polypeptide synthesis (Jacob and Monod, 1961). The enzyme responsible for the synthesis of RNA couples together the ribonucleoside triphosphates ATP, GTP, CTP and UTP on DNA templates in a reaction that is driven by the release and subsequent hydrolysis of pyrophosphate group. In addition to the mRNA, a transcript of structural genes, two other main types of RNA exist in all cells: ribosomal RNA (rRNA), responsible for the structural integrity of ribosomes as well as their catalytic function in protein synthesis; and transfer RNA (tRNA), which provides activated amino acids to the site of the polypeptide chain synthesis.

1.2 RNA polymerases

Prokaryotic cells are known to contain one RNA polymerase that directs the synthesis of all cell's RNA except short RNA primers used in the process of DNA replication. Eukaryotic cells contain three types of RNA polymerases that differ in the RNAs they synthesize: RNA polymerase I, located in the nucleoli, which synthesizes precursors of rRNAs; RNA polymerase II, located in the nucleoplasm, which synthesizes mRNA and microRNA precursors (Lee et al, 2004); and RNA polymerase III, located in the nucleoplasm, which synthesizes precursors of 5S rRNA, tRNAs and a variety of other small RNAs. In addition to these three types of polymerases, eukaryotic cells contain special mitochondrial and (in plants) chloroplast RNA polymerases that direct the transcription of organellar genes (Voet and Voet, 1995).

RNA polymerases are large proteins that contain between 5 and 15 subunits with total molecular weight up to 0.6 MDa. Recent X-ray crystallographic studies revealed structures of bacterial protein from *Thermus aquaticus* at 3.3 Å resolution (Zhang et al, 1999) and RNA polymerase II from the yeast *Saccharomyces cerevisiae* at 2.8 Å resolution (Cramer et al, 2001). These structures now form a basis for understanding the function of all RNA polymerases and dissecting the transcription mechanism by site-directed mutagenesis and further structural studies. Figure 1.2 shows the general architecture of RNA polymerases. All 10 subunits in the yeast RNA polymerase II structure are identical or homologous to subunits of RNA polymerases I and III (Geiduschek and Bartlett, 2000). RNA polymerase II is also highly conserved across species. Yeast and human RNA polymerase II sequences exhibit 53% overall identity, and the conserved residues are distributed over the entire structure (Cramer et al, 2001). Sequence conservation between yeast and bacterial RNA polymerases is far smaller than for yeast and human enzymes, but the structural homology of these two proteins is very extensive (Sweetser et al, 1987; Zhang et al, 1999). They share the core structure, and thus a conserved catalytic mechanism, but differ entirely in peripheral and surface structure where interactions with other proteins (such as general transcription factors and regulatory factors) take place.

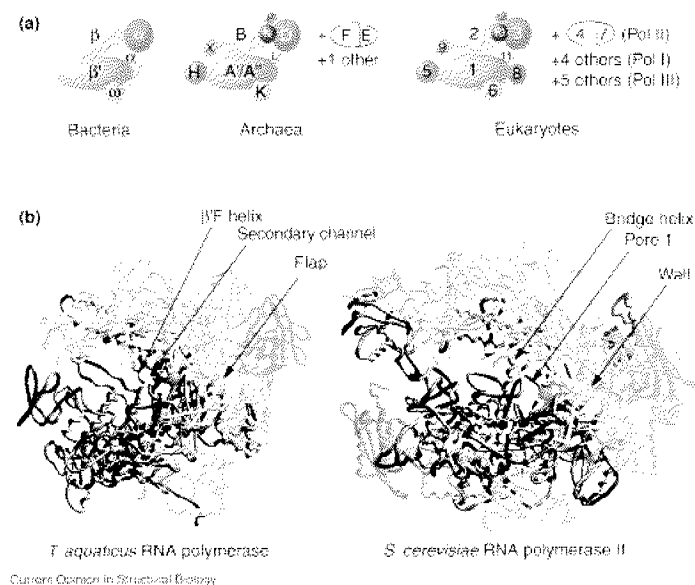


Figure 1.2. RNA polymerase subunit architecture. **(a)** Schematic presentation and color code for RNAP subunits in bacterial, archaeal and eukaryotic RNA polymerases. Corresponding subunits have the same color. **(b)** Three-dimensional structures of *T. aquaticus* RNA polymerase and *S. cerevisiae* RNA polymerase II shown in the same orientation. The subunits are colored according to (a). The active site metal ion A is shown as a pink sphere. Zinc ions are shown as blue spheres. (Cramer, 2002).

The crystal structure of an actively transcribing complex of the yeast RNA polymerase II with DNA at 3.3 Å resolution provided insights into the mechanistic function of the enzyme (Gnatt et al, 2001). Figure 1.3 shows the RNA polymerase II transcribing complex, indicating the paths of the nucleic acids and the locations of some functional elements of the enzyme. DNA entering the enzyme is gripped by protein "jaws". The 3' (growing) end of the RNA is located adjacent to an active site Mg^{2+} ion. A "wall" of protein blocks the straight passage of polynucleic acids through the enzyme, as a result of which the axis of the DNA-RNA makes almost a right angle with the axis of the entering DNA. The bend exposes the end of the DNA-RNA hybrid for addition of substrate nucleoside triphosphates (NTPs). The NTPs may enter through a funnel-shaped opening on the underside of the enzyme and gain access to the active center through a pore. The 5' end of the RNA abuts a loop of protein (the rudder), which prevents extension of the DNA-RNA hybrid beyond 9 base pairs, separating DNA from RNA. The exit path of the RNA passes beneath the rudder and beneath another loop of protein (the lid) (Klug, 2001).

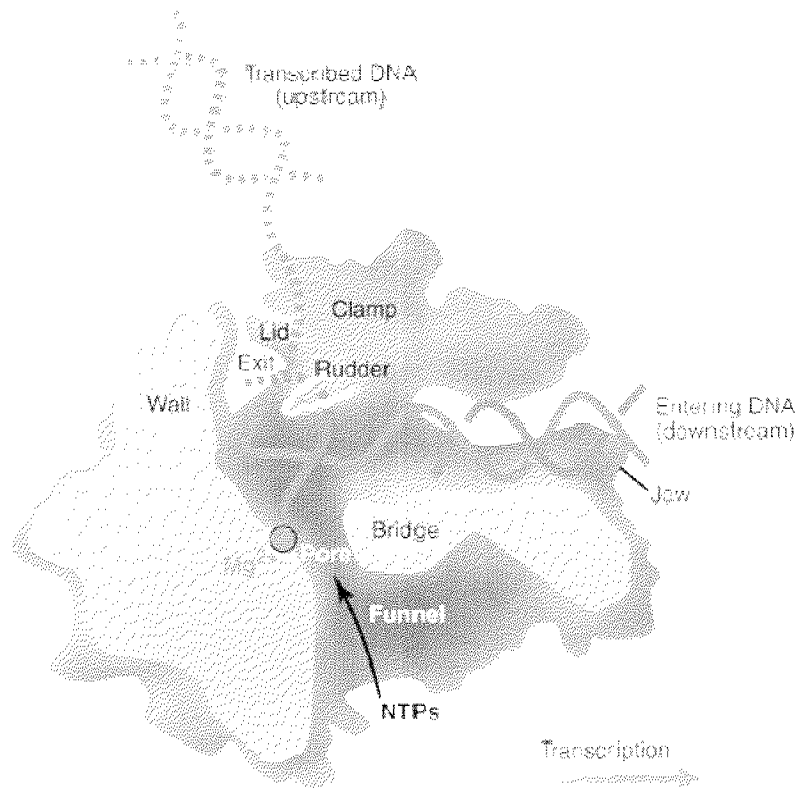


Figure 1.3. RNA polymerase II transcribing complex.

Cut surfaces of the protein, in the front, are lightly shaded. By convention, the polymerase is moving on the DNA from left to right. The template DNA strand is in blue, the nontemplate strand is in green, and the RNA in the DNA-RNA hybrid in the active center region is in red. (Klug, 2001).

1.3 RNA polymerase II core promoter

1.3.1 Initiation of transcription in Eukaryotes

Most eukaryotic organisms, including humans, have thousands of genes. Expression of each of them is well regulated in order for the organism to survive, grow and develop. There is a wide variety of regulatory mechanisms acting on DNA, RNA or protein level that control gene expression. Regulation of RNA polymerase transcription is one of the most important ones, as this process is a first step in the flow of genetic information.

1.3.2 Core promoter elements

RNA polymerase II core promoter is the most important element in the control of transcription initiation and the ultimate target of action of many factors that are involved in the regulation of transcription. The core promoter includes DNA elements that can extend ~35 bp upstream and/or downstream of the transcription initiation site. For RNA polymerase II to accurately and efficiently initiate transcription from the core promoter, it requires additional factors that are commonly termed the "basal" or "general" transcription factors (GTFs). These include the transcription factor (TF) IIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIF (Butler and Kadonaga, 2002).

Figure 1.4 shows several sequence motifs that are commonly found in core promoters. These include the TATA box, initiator (Inr), TFIIB recognition element (BRE), and downstream core promoter element (DPE). Each of these core promoter elements is found in some but not all core promoters, as there are no universal core promoter elements (Smale and Kadonaga, 2003).

In addition to the core promoter, there are other *cis*-acting DNA sequences that can regulate RNA polymerase II transcription. These elements include the proximal promoter, enhancers, silencers, and boundary/insulator elements and contain recognition sites for a variety of sequence-specific DNA-binding factors (Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999). The proximal promoter is the region in the immediate vicinity of the transcription start site (roughly from -250 to +250), while enhancers and silencers can be located many kbp from the transcription start site. Boundary/insulator elements appear to prevent

spreading of the activating effects of enhancers or the repressive effects of silencers or heterochromatin (West et al. 2002).

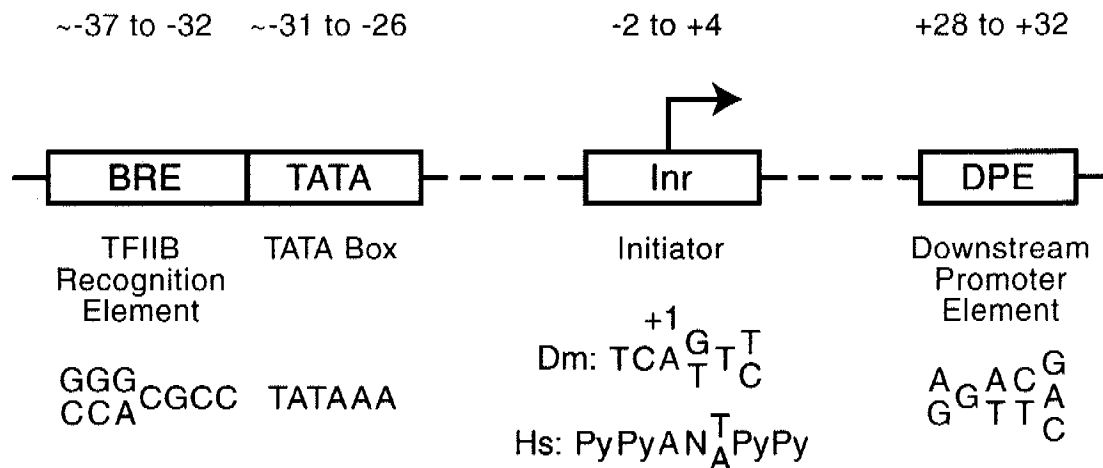


Figure 1.4. RNA polymerase II core promoter elements.

Any specific core promoter may contain some, all, or none of these motifs. The BRE is an upstream extension of a subset of TATA boxes. The DPE requires an Inr, and is located precisely at +28 to +32 relative to the A₊₁ nucleotide in the Inr. The DPE consensus was determined with *Drosophila* transcription factors and core promoters. The Inr consensus sequence is shown for both *Drosophila* (Dm) and humans (Hs). (Butler and Kadonaga, 2002).

A stable preinitiation complex (PIC) can form *in vitro* on TATA-dependent core promoters by association of the basal factors in the following order: TFIID/TFIIA, TFIIB, RNA polymerase II/TFIIF, TFIIE, and then TFIIH (Orphanides et al, 1996; Lee and Young, 2000; Woychik and Hampsey, 2002). In the stepwise PIC assembly mechanism described above, TFIID and TFIIB are the first two factors that interact with the core promoter. Accordingly, it appears that these two factors have a critical role in the recognition of core promoter motifs. TFIID is a multisubunit protein that consists of TBP (the TATA box-binding protein) and up to 14 TBP-associated factors (TAFs) (Burley and Roeder, 1996; Albright and Tjian, 2000; Verrijzer, 2001). TFIIB is a single polypeptide that interacts with TBP as well as with the DNA upstream of the TATA box (Nikolov et al, 1995; Tsai and Sigler, 2002).

1.3.2.1 The TATA box

The TATA box was the first core promoter element discovered in eukaryotic protein-coding genes (Goldberg, 1979; Breathnach and Chambon, 1981). It was identified from a comparison of the 5' flanking sequences in a number of *Drosophila*, mammalian, and viral protein-coding genes. The TATA box consensus sequence TATAAA is located 25 to 30 bp upstream of the transcription start site. In *Saccharomyces cerevisiae*, TATA boxes were also found to be critical for transcription initiation; but in this organism, the element was located 40–120 bp from the start site (Struhl, 1989). Mutations in the TATA box usually reduce or abolish the activity of cellular and viral promoters (Breathnach and Chambon, 1981; Grosschedl and Birnstiel, 1980; Wasylyk et al, 1980; Hu and Manley, 1981). However, a wide range of sequences can function as a TATA box in vivo (Singer et al. 1990). As for the other core promoter elements, TATA boxes are found in a subset of core promoters. In humans, it was found that only about 32% of 1031 potential promoter regions contain a putative TATA box motif (Suzuki et al. 2001).

Even though TBP is the predominant TATA box-binding protein, other TBP-related factors (TRFs) were found in humans and *Drosophila* (Berk, 2000). *Drosophila* TRF1 can bind TATA box motifs (Crowley et al, 1993; Hansen et al, 1997), but it also binds a motif termed the TC box, which is not bound efficiently by TBP (Holmes and Tjian 2000). These and other data suggest that TRF1 mediates transcription in a complex that is distinct from TFIID at TC box-containing promoters (Holmes and Tjian 2000). Moreover, in *Drosophila*, transcription by RNA polymerase III involves TRF1, rather than TBP as in other organisms such as yeast and humans (Takada et al. 2000). *Drosophila* and human TRF2 generally do not appear to bind to TATA box motifs, but they are required for the expression of a specific set of genes (Dantonel et al. 2000; Kaltenbach et al. 2000).

In terms of transcription directionality, TBP does not seem to be sufficient for its determination, even though it binds the asymmetric TATA box sequence in a polar manner in TBP-DNA cocrystals (Kim et al, 1993). The polar binding of TBP can, in theory, lead to the assembly of a properly oriented preinitiation complex and influence the direction of transcription. In the context of more complex synthetic promoters or native promoters, the main determinants of directionality appear to be the relative locations of the activator binding sites,

TATA box, and other core promoter elements (Xu et al, 1991; O'Shea-Greenfield and Smale, 1992).

The mechanism that determines the distance from the TATA box to the transcription start site has not yet been clarified. TFIIB and the RNA polymerase itself have been implicated in this process (Fairley et al, 2002; Berroteran et al, 1994). A recent extensive study of sequenced mouse and human genomes has shown that generally, promoters can be divided into 2 classes. One is the conserved, tissue-specific, TATA-box-containing class, with a single transcription start and the other is an evolvable, CpG rich class (described in section 1.3.2.5) with more than one transcription start and which represents majority of the promoters (Carninci et al, 2006).

1.3.2.2 The initiator (Inr)

The initiator element (Inr) was defined as a discrete core promoter element, functionally similar to the TATA box, that can function independently of a TATA box in an analysis of the lymphocyte-specific terminal transferase (TdT) promoter (Smale and Baltimore, 1989). An extensive mutant analysis revealed that the sequence between -3 and $+5$ is necessary and sufficient for accurate transcription *in vitro* and *in vivo* (Smale and Baltimore, 1989; Javahery et al, 1994). This region matched the start site consensus sequence observed in comparisons of promoter sequences from efficiently transcribed protein-coding genes. These comparisons revealed that most contained an adenosine at the transcription start site ($+1$), a cytosine at the -1 position, and a few pyrimidines surrounding these nucleotides (Corden et al, 1980).

Inr elements are found in both TATA-containing as well as TATA-less core promoters. The consensus for the Inr in mammalian cells is Py-Py(C)-A₊₁-N-T/A-Py-Py (Corden et al. 1980; Bucher 1990; Javahery et al. 1994; Lo and Smale 1996). In *Drosophila*, the Inr consensus is T-C-A₊₁-G/T-T-C/T (Hultmark et al. 1986; Purnell et al. 1994; Arkhipova 1995; Kutach and Kadonaga 2000). The A₊₁ position is designated at the $+1$ start site because transcription commonly initiates at this nucleotide. More generally, however, transcription initiates at a single site or in a cluster of multiple sites in the vicinity of the Inr (and not necessarily at the A₊₁ position).

A variety of factors have been found to interact with the Inr element. There is evidence that TFIID binds to the Inr in a sequence-specific manner (Kaufmann and Smale 1994; Martinez

et al. 1994; Purnell et al. 1994; Burke and Kadonaga 1996; Oelgeschläger et al. 1996). TAF1 and TAF2 subunits of TFIID have been suggested as the key subunits for this interaction (Verrijzer et al. 1994, 1995; Kaufmann et al. 1998; Chalkley and Verrijzer 1999). The domains of TAF1 and TAF2 that are responsible for Inr recognition have not been determined. However, functional studies have shown that a trimeric TBP-TAF1-TAF2 complex is sufficient for Inr activity in reconstituted transcription assays (Verrijzer et al, 1995). TAF2 was also implicated in Inr activity by the finding that *Drosophila* TAF2 possesses a core promoter-binding activity (Verrijzer et al, 1994) and that human TAF2 corresponds to a biochemical activity in HeLa cell extracts that is required for Inr function (Kaufmann et al, 1996; Kaufmann et al, 1998).

The synergistic function of TATA and Inr elements was found to correlate with the cooperative binding of TFIID to the two elements. The general transcription factor TFIIA is reported to be critical for the cooperative binding of TFIID to the Inr element (Emami et al, 1997). This observation is consistent with earlier evidence that TFIIA can induce a conformational change in the TFIID complex, which alters its contacts with DNA in the vicinity and downstream of the transcription start site (Horikoshi et al, 1988; Chi and Carey, 1996). In addition, crosslinking of TAF1 to the Inr of the AdML promoter is greatly enhanced in the presence of TFIIA, probably due to a TFIID conformational change (Oelgeschlager et al, 1996).

Interestingly, purified RNA polymerase II (or RNA polymerase II along with TBP, TFIIB, TFIIF) is also able to recognize the Inr and to mediate transcription in an Inr-dependent manner in the absence of TAFs (Carcamo et al. 1991; Weis and Reinberg 1997). Taken together, these results suggest that TFIID and RNA polymerase II may recognize and interact with the Inr at different steps in the transcription process.

There are other sequence-specific DNA-binding proteins that have been found to interact with the Inr and may participate in Inr-dependent transcription at a subset of promoters: TFII-I and YY1. TFII-I is a basic-helix-loop-helix (bHLH) protein that binds to Inr and E-box (CACGTG) elements and stimulates transcription *in vitro*, interacts with other sequence-specific DNA-binding factors and may thus participate in the communication between promoter-binding factors and the basal transcriptional machinery (Roy et al. 1991, 1997; Cheriyaath et al. 1998; Grueneberg et al. 1997). YY1 is a zinc finger protein that binds to the Inr motifs in the AAV P5 and human DNA polymerase β core promoters (Usheva and Shenk 1994; Weis and Reinberg

1997). With the AAV P5 promoter (as a supercoiled DNA template), transcription was observed with purified YY1, TFIIB, and RNA polymerase II (Usheva and Shenk 1994).

1.3.2.3 The downstream core promoter element (DPE)

The downstream core promoter (DPE) was identified as a motif required for the binding of purified TFIID to a subset of TATA-less promoters (Butler and Kadonaga, 2002). The DPE is conserved from *Drosophila* to humans and is typically but not exclusively found in TATA-less promoters. It was shown that the DPE acts in conjunction with the Inr, with its core sequence (Figure 1.4) located at precisely +28 to +32 relative to the A₊₁ nucleotide in the Inr motif (Kutach and Kadonaga, 2000).

TFIID binds cooperatively to the DPE and Inr motifs as one unit, with a strict spacing requirement (Burke and Kadonaga, 1996; Kutach and Kadonaga, 2000). In this respect, the DPE differs from the TATA box, which is able to function independently of the presence of an Inr.

Photocrosslinking studies with purified TFIID indicated that TAF6 and TAF9 are in close proximity to the DPE (Burke and Kadonaga, 1997). Genetic analysis of TAF6 and TAF9 carried out in *Drosophila* reveal them as essential genes. Also, TAF6 and TAF9 mutations have been found that can increase or decrease the expression of DPE-containing genes, but it is not yet known whether these mutations affect the function of TFIID at DPE-dependent promoters (Soldatov et al, 1999; Aoyagi and Wassarman, 2001).

1.3.2.4 The TFIIB recognition element (BRE)

The TFIIB recognition element (BRE) is a TFIIB binding site located immediately upstream of some TATA boxes (Lagrange et al. 1998). The interaction of TFIIB with the BRE was analyzed in detail by X-ray crystallography of a TFIIB-TBP-DNA complex (Tsai and Sigler 2000). The BRE consensus is G/C-G/C-G/A-C-G-C-C (where the 3' C of the BRE is followed by the 5' T of the TATA box), and at least a 5 out of 7 match with the BRE consensus was found in 12% of a collection of 315 TATA-containing promoters (Lagrange et al. 1998). *In vitro* transcription experiments with purified general transcription factors revealed that the BRE facilitates the incorporation of TFIIB into the productive transcription initiation complexes (Lagrange et al. 1998). Other studies (*in vitro* transcription with a crude extract or by transient transfection analysis) showed that the BRE has a negative effect on transcription (Evans et al.

2001). Additional experiments suggested that the Gal4-VP16 activator can increase transcription, at least in part, by disruption of the repressive TFIIB-BRE interaction (Evans et al. 2001). These results, which are not necessarily contradictory, show that BRE might function as a positive or negative core promoter element.

1.3.2.5 CpG islands

The CpG dinucleotide is underrepresented in vertebrate genomes due to methylation at the position 5 of the cytosine ring and subsequent deamination of the 5-methylcytosine to give a TpG dinucleotide that is not repaired by the DNA repair machinery. However, there are stretches of DNA, termed CpG islands, that are relatively GC-rich and overrepresented in CpG dinucleotides that are mostly unmethylated (Bird 1986; Adachi and Lieber 2002).

CpG islands range in size from 0.5 to 2 kbp and contain promoters for a wide variety of genes. They lack TATA or DPE core promoter elements, but contain multiple GC box motifs that are bound by Sp1 and related transcription factors (Brandeis et al. 1994; Macleod et al. 1994). Transcription from CpG islands initiates from multiple weak start sites that are often distributed over a region of about 100 bp, in contrast to transcription from TATA or DPE-dependent core promoters that occurs from a single site or localized cluster of sites. The alternative transcripts resulting from CpG islands occur in 58% of protein coding genes and generate truncated proteins or eliminate the protein product (Carninci et al, 2006). The molecular determinants of transcription from CpG islands are yet to be identified. Sp1 transcription factor has been implicated to act in conjunction with an Inr motif to generate the array of start sites that are observed (Smale and Baltimore, 1989; Emami et al. 1995).

1.3.3 Transcriptional regulation via core promoter elements

The core promoter is the ultimate target of sequence-specific transcription factors and co-regulators. Current evidence indicates that many transcriptional enhancers and factors exhibit core promoter specificity. Hence, core promoters are not passive elements that serve only to direct the proper placement of the RNA polymerase II transcriptional machinery, but also *cis*-acting regulatory elements that provide another level of transcriptional regulation. Importantly, core-promoter-specific enhancers were identified which act in conjunction with their cognate core promoters. For example, if a DPE-specific enhancer was fused to a TATA-dependent core

promoter, then transcriptional activation would not be observed (Butler and Kadonaga, 2002). Therefore, transcriptional regulation has been found to be a process of substantial complexity and diversity.

1.4 General transcription factor TFIID

Transcription factor IID (TFIID) was first identified by Roeder and colleagues as a biochemical activity that elutes from a phosphocellulose column between 0.6 and 1 M NaCl. This activity was essential for the activity of TATA-containing core promoters and was capable of binding the core promoter from the adenovirus major late promoter in a DNase I footprinting assay (Sawadogo and Roeder, 1985; Matsui et al, 1980). However, further purification and cloning of TFIID proved to be unusually difficult because, in *Drosophila* and man, it appeared to be a heterogeneous multiprotein complex (Nakajima et al, 1988). The breakthrough was made when yeast TBP was isolated (Cavalini et al, 1989), followed by the identification of human and *Drosophila* homologs. Using anti-TBP antibodies to purify associated proteins from *Drosophila* and human cells, several polypeptides designated TAFs (TBP-associated factors) were identified (Dymlacht et al., 1991; Tanese et al., 1991). It was found later that TBP actually participates in transcription initiation of all three RNA polymerases and that the associated proteins (TAFs) confer their specificity for each polymerase system on the TBP-containing complexes (Hernandez, 1993).

1.4.1 TBP-associated factors (TAFs)

Subunits of TFIID (TAFs) were identified soon after the purification of the complex (Albright and Tjian, 2000). The amino acid sequences of many TAFs revealed a high level of conservation. TFIID structure has changed little through evolution, with all species containing at least 13 conserved subunits (Gangloff et al., 2001). Recently, a consensus on the nomenclature of TAFs has been developed to avoid confusion (Tora, 2002). The currently identified TAFs are listed in Table 1.1 (Matangkasombut et al., 2004).

New name	<i>Homo sapiens</i>	<i>Drosophila melanogaster</i>	<i>Caenorhabditis elegans</i> (Previous name)	<i>Ce</i> New name	<i>Saccharomyces cerevisiae</i>	<i>Saccharomyces pombe</i>
TAF1	TAF _{II} 250	TAF _{II} 230	<i>taf-1</i> (W04A8.7)	<i>taf-1</i>	Taf145/130p	TAF _{II} 111
TAF2	TAF _{II} 150	TAF _{II} 150	<i>taf-2</i> (Y37E11B.4)	<i>taf-2</i>	Taf150p or TSM1	(T38673)
TAF3	TAF _{II} 140	TAF _{II} 155 or BIP2	(C11G6.1)	<i>taf-3</i>	Taf47p	
TAF4	TAF _{II} 130/135	TAF _{II} 110	<i>taf-5</i> (R119.6)	<i>taf-4</i>	Taf48p or MPT1	(T50183)
TAF4B	TAF _{II} 105					
TAF5	TAF _{II} 100	TAF _{II} 80	<i>taf-4</i> (F30F8.8)	<i>taf-5</i>	*Taf90p	TAF _{II} 72
TAF5B						TAF _{II} 73
TAF5L	*PAF65β	Cannonball				
TAF6	TAF _{II} 80	TAF _{II} 60	<i>taf-3.1</i> (W09B6.2)	<i>taf-6.1</i>	*Taf60p	(CAA20756)
TAF6L	*PAF65α	(AAF52013)	<i>taf-3.2</i> (Y37E11A.8)	<i>taf-6.2</i>		
TAF7	TAF _{II} 55	(AAF54162)	<i>taf-8.1</i> (F54F7.1)	<i>taf-7.1</i>	Taf67p	TAF _{II} 62/PTR6
TAF7L	TAF2Q		<i>taf-8.2</i> (Y111B2A.16)	<i>taf-7.2</i>		
TAF8	(BAB71460)	Prodos	(ZK1320.7)	<i>taf-8</i>	Taf65p	(T40895)
TAF9	*TAF _{II} 32/31	TAF _{II} 40	<i>taf-10</i> (T12D8.7)	<i>taf-9</i>	*Taf17p	(S62536)
TAF10	*TAF _{II} 30	TAF _{II} 24	<i>taf-11</i> (K03B4.3)	<i>taf-10</i>	*Taf25p	(T39928)
TAF10B		TAF _{II} 16				
TAF11	TAF _{II} 28	TAF _{II} 30β	<i>taf-7.1</i> (F48D6.1)	<i>taf-11.1</i>	Taf40p	(CAA93543)
TAF11L			<i>taf-7.2</i> (K10D3.3)	<i>taf-11.2</i>		
TAF12	*TAF _{II} 20/15	TAF _{II} 30α	<i>taf-9</i> (Y56A4.3)	<i>taf-12</i>	*Taf61/68p	(T37702)
TAF13	TAF _{II} 18	(AAF53875)	<i>taf-6</i> (C14A4.10)	<i>taf-13</i>	Taf19p or FUN81	(CAA19300)
TAF14					TAF30	
TAF15	TAF _{II} 68					
			B-TFIID			
BTAF1	TAF _{II} 170/TAF-172	Hcl89B			Mot1	

Table 1.1. TAF nomenclature including the corresponding known orthologues and paralogues.
(Taken from Matangkasombut et al., 2004.)

1.4.2 TFIID architecture

Structural studies of TBP and TAF subcomplexes provided valuable insight into the mechanisms of TFIID function and complex organization. Figure 1.5 shows a low resolution structure of the whole human TFIID complex obtained by electron microscopy (Andel et al., 1999). The complex has a horseshoe shape, consisting of 3–4 lobes connected by flexible regions. It has been suggested that the concave face may represent the DNA binding surface and that DNA might be accommodated in the central channel. TBP was mapped to the central lobe, which is also where TFIIA and TFIIB bind.

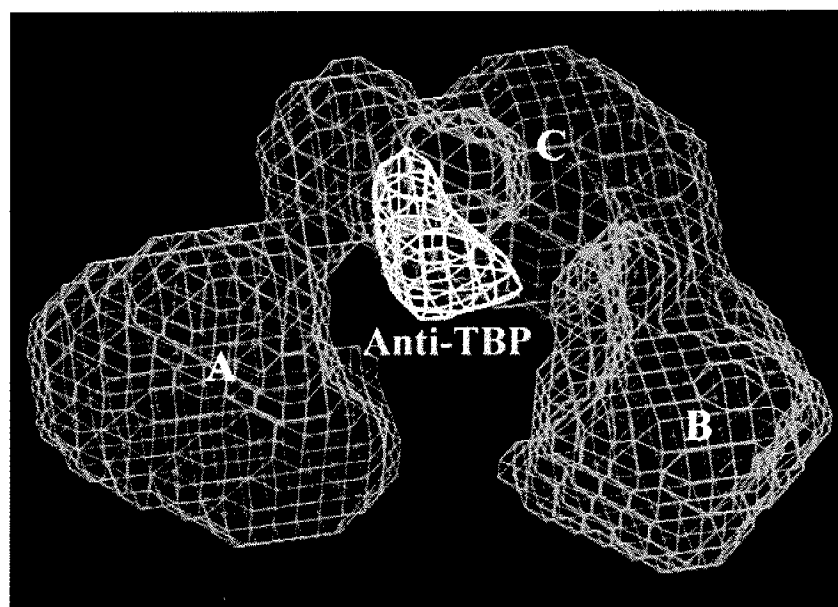


Figure 1.5. The EM structure of the human TFIID complex at low resolution. Three lobes are marked “A”, “B” and “C”. The position of the TBP is shown in yellow.

TFIID is formed through numerous protein–protein interactions, both TBP–TAF and TAF–TAF (Burley and Roeder, 1996; Gangloff et al., 2001). Many interactions were identified *in vitro*, and some have been verified *in vivo* by genetic studies in yeast. It has initially been suggested that the largest TAF subunit, TAF1, is a scaffold for the complex assembly because it interacts directly with TBP and with several other TAFs (Chen et al, 1996). However, recent *in*

1.4.3 Other TAF-containing complexes

TAFs are not restricted to TFIID; a subset of TAFs is found in several other transcription-related complexes, such as the histone acetyltransferase (HAT) complexes in yeast and humans and the Polycomb group complex in *Drosophila*. *In vivo* depletion of these shared TAFs results in broad defects in RNA polymerase II transcription. Examples of such complexes are yeast SAGA and mammalian PCAF. SAGA is known to be required for nucleosome acetylation and transcriptional stimulation (Grant et al, 1998). PCAF is a histone acetylase complex of about 20 subunits that has a role in regulation of transcription, cell cycle progression, and differentiation (Ogryzko et al, 1998).

There are also the alternative TFIID complexes found in multicellular eukaryotes, which are tissue-specific or developmental stage-specific forms of TFIID. These complexes contain either a variant of TBP or variant TAF subunits. A TBP-free TAF-complex (TFTC) has also been described, which is able to direct the formation of the pre-initiation complex (Wieczorek et al, 1998).

1.5 *Outline of the thesis*

TFIID is a key factor in regulation of RNA polymerase II transcription initiation. It is involved not only in core promoter recognition, but also acts as a partner of transcription activators/repressors and possesses enzymatic activities that are thought to aid in transcription activation. However, very little high resolution structural information is available on this important complex, due to the technical difficulties in obtaining sufficient quantities of pure material.

The focus of this work is human TAF2, the second largest subunit of human TFIID which is implicated in the recognition of initiator core promoter element together with hTAF1. Chapter 3 describes the development of the purification protocol of hTAF2 recombinantly produced in Sf21 insect cells. This protocol provided material of high purity, enabling extensive crystallization trials that were conducted during this work. By combination of the limited proteolysis and gel filtration studies, described in the Chapter 4, hTAF2 was shown to consist of two structural domains. One is a large globular domain of compact structure and 116 kDa in size, representing the structural core of hTAF2. The other is the smaller, 22 kDa structurally disordered domain, termed “TAF2 tail”. These studies have provided guidelines for the future design of hTAF2 crystallization constructs. Furthermore, functional analysis has been performed in order to elucidate the role of hTAF2 in core promoter recognition. These studies, described in the Chapter 5, reveal that hTAF2 has a non-specific DNA binding activity, localized to the lysine-rich tail domain. To obtain more information on the interactions hTAF2 forms with other subunits of TFIID, in vitro binding experiments have been conducted with human TAF8/TAF10 complex. These experiments show that hTAF2 directly interacts with hTAF8/TAF10, revealing important new information about the architecture of TFIID.

2

MATERIALS AND METHODS

2.1 Materials, chemicals and equipment

2.1.1 General chemicals

Chemical	Manufacturer	Number
1,4-Dithiothreitol (DTT)	Fluka	43819
2'-Deoxyadenosine 5'-triphosphate (dATP)	Pharmacia	27-1850-03
2'-Deoxycytidine 5'-triphosphate (dCTP)	Pharmacia	27-1860-03
2'-Deoxyguanosine 5'-triphosphate (dGTP)	Pharmacia	27-1870-03
2'-Deoxythymidine 5'-triphosphate (dTTP)	Pharmacia	27-1880-03
2-Mercaptoethanol (BME)	Fluka	63690
2-Morpholindethansulfonic acid (MES)	Fluka	69890
4-(2-Hydroxyethyl)-piperazine-1-ethane-sulfonic acid (HEPES)	Fluka	54459
Acetonitril	Fisher Scientific	A/0627/17
Acrylamide	Fluka	01699
Acrylamide 4K – Solution (30%) Mix 37.5:1	AppliChem	A1672,1000
Adenosine triphosphate (ATP)	Boehringer Mannheim	519979
Adenosine triphosphate (ATP)	Fluka	02060
Agar granulated	Difco	0145-17-0
Agarose, electrophoresis grade	Invitrogen	15510-027
Amberlite MB3	BDH	55008
Ammonium peroxydisulphate (AMPS)	Merk	708K3624901
Ammonium sulfate	Fluka	09982
Ampicillin	Chemie Brunswick	018040
Bacto tryptone	Difco	0123-17-3
Benzamidine	Fluka	12073
Betaine monohydrate	Fluka	14300
Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)-methane (Bis-Tris)	Fluka	14880
Boric acid	Fluka	15660
Bovine serum albumin (BSA)	Boehringer Mannheim	238031
Bromophenol blue	Fluka	18030
Cacodylic acid sodium salt	Fluka	20840
Calcium acetate	Fluka	21056
Chloramphenicol	Fluka	23275
Chloroform	Fluka	25690
Citric acid monohydrate	Fluka	27488
Coomassie brilliant blue R 250	Fluka	27815
dNTP Mix, 10 mM each	Fermentas	R0191
D-sorbitol	Sigma	S-1876
Deoxyribonucleic acid lambda (λ DNA)	MBI Fermentas	SD00 1
Dimethylsulphoxide (DMSO)	Fluka	41640
Di-Potassium hydrogen phosphate, microselect	Fluka	60354
Ethanol for HPLC	Fluka	02855
Ethanolamine	Fluka	02400
Ethidiumbromide	Sigma	E8751
Ethylendiamine tetraacetic acid disodium salt dihydrate (EDTA)	Fluka	03679
Ethylenglycol-O,O'-bis(2-aminoethyl)-N,N,N',N'- tetraacetic acid	Sigma	E4378
Gentamycin sulphate	Fluka	48760
Glycerol anhydrous	Fluka	49770
Glycine	Serva	23390
Guanidine hydrochloride (GdnHCl)	Fluka	50940
Hydrochloric acid (HCl)	Fluka	29345
Imidazole	Fluka	52834
Isopropanol	Merck	9634
Isopropyl- β -D-thiogalactoside (IPTG)	Bachem	Q-1280
Kanamycin sulphate	Fluka	60615
L-leucine p-nitroanilide	Sigma-Aldrich	L 9125

L-alanine 4-nitroanilide	Sigma-Aldrich	A 9325
L-lysine p-nitroanilide dihydrobromide	Sigma-Aldrich	L 7002
Magnesium acetate	Fluka	63051
Magnesium chloride	Fluka	63064
Magnesium nitrate	Fluka	63084
Magnesium sulfate	Fluka	63140
Methanol	Fluka	65543
Methylene blue	Fluka	66721
NBT/BCIP Stock Solution (18.75 mg/ml Nitro blue tetrazolium chloride and 9.4 mg/ml 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt in 67% (v/v) DMSO)	Roche	1 681 451
N,N,N',N'-tetramethylethylenediamine (TEMED)	Fluka	87689
N,N'-methylene bisacrylamide (bisacrylamide)	Fluka	66669
Nitric acid 65%	Fluka	84380
Nonidet P40	Fluka	74385
Peptstatin A	Fluka	77170
Phenol	Fluka	77613
Phenylmethanesulfonyl fluorided (PMSF)	Fluka	78830
Poly(dG-dC)- Poly(dG-dC)	Amersham Biosciences	27-7910-01
Poly(dI-dC)- Poly(dI-dC)	Amersham Biosciences	27-7880-01
Polyethylene glycol 6000 (PEG 6000)	Fluka	81253
Polyethylene glycol 8000 (PEG 8000)	Fluka	81268
Ponceau S	Fluka	81460
Potassium acetate	Fluka	60034
Potassium chloride	Fluka	60130
Potassium dihydrogen phosphate	Fluka	60230
Potassium fluoride	Fluka	60240
Potassium nitrate	Fluka	60419
sec Butanol	Fluka	19440
Silicon oil DC 200, 110mPa.s	Fluka	85414
Sodium dodecyl sulfate (SDS)	Fluka	71729
Sodium acetate	Fluka	71180
Sodium azide	Fluka	71290
Sodium chloride	Fluka	71379
Sodium dihydrogenphosphate Dihydrate	Fluka	71645
Sodium hydroxide	Merck	6498
Sodium nitrate	Fluka	71758
Streptomycine sulphate	Fluka	85880
Sulfuric acid	Fluka	28347
Tetracycline hydrochloride	Fluka	87130
Tricine	Merck	8602
Triethanolamine	Fluka	90279
Triton X-100	Fluka	93420
Trizma base (Tris)	Sigma	T1503
Tween-20	BioRad	170-6531
Urea	BDH	102908D
X-gal	Invitrogen	15520-018
Xylene cyanole	Fluka	95600
Yeast Extract	Difco	0596
Zinc acetate	Merck	8802
Zinc chloride	Fluka	96470

2.1.2 Buffers and solutions

Buffers and solutions	Composition
1x <i>Bam</i> H I buffer (NEB #136L)	10 mM Tris-HCl pH 7.9 150 mM NaCl 1 mM DTT
1x <i>Eco</i> R I buffer (NEB #101L)	100 mM Tris-HCl pH 7.5 10 mM MgCl ₂ 50 mM NaCl 0.025% Triton X-100

1x <i>EcoR</i> V buffer	50 mM Tris-HCl pH 7.9 10 mM MgCl ₂ 100 mM NaCl 1 mM DTT
1x NEB buffer #1 (NEB #007-1)	10 mM BisTris Propane-HCl pH 7.0 10 mM MgCl ₂ 1 mM DTT
1x NEB buffer #2 (NEB #007-2)	10 mM Tris-HCl pH 7.9 10 mM MgCl ₂ 50 mM NaCl 1 mM DTT
1x NEB buffer #3 (NEB #007-3)	50 mM Tris-HCl pH 7.9 10 mM MgCl ₂ 100 mM NaCl 1 mM DTT
1x NEB buffer #4 (NEB #007-4)	20 mM Tris-acetic acid pH 7.9 10 mM MgOAc ₂ 50 mM KOAc 1 mM DTT
1x TAE	40 mM Tris-acetic acid pH 8.3 40 mM acetic acid 1.0 mM EDTA
1x TBE	89 mM Tris-boric acid pH 8.3 89 mM boric acid 2.5 mM EDTA
1x TBS	25 mM Tris-HCl pH 8.0 150 mM NaCl
10x CIP buffer (Roche, 1 243 284)	0.5 M Tris-HCl pH 9.3 10 mM MgCl ₂ 1 mM ZnCl ₂ 10 mM spermidine
10x Pfu buffer	200mM Tris-HCl (pH 8.8 at 25°C) 100mM (NH ₄) ₂ SO ₄ 100mM KCl 1% (v/v) Triton X-100 1mg/ml BSA
6 x DNA gel loading buffer	2.5 mg/ml bromophenol blue 2.5 mg/ml xylene cyanol 30% (v/v) glycerol 60 mM EDTA-NaOH pH 8.0
Acrylamide 30% (60:1)	75 g acrylamide 1.25 g bisacrylamide water to 250 ml
Ampicilin stock solution (100 mg/ml)	1 g ampicilin water to 10 ml, filter sterilized
AMPS 25%	2.5 g AMPS water to 10 ml, filter sterilized
Bacterial lysis buffer	50 mM Tris-HCl pH 7.5 100 mM NaCl
Blot destain	45% (v/v) absolute ethanol 7% (v/v) acetic acid
Blot stain	0.1% (w/v) amido black 40% (v/v) methanol (HPLC grade) 1% (v/v) acetic acid
BSA 10x	10 mg/ml BSA acetylated (in water)
Chloramphenicol stock solution (34 mg/ml)	0.34 g chloramphenicol absolute ethanol to 10 ml
CIA	96% (v/v) chloroform 4% (v/v) isoamyl alcohol
Gentamycine stock solution (7 mg/ml)	0.35 g gentamycine water to 50 ml, filter sterilized
Insect cell lysis buffer	50 mM Tris-HCl pH 7.4 100 mM KCl 0.1% NP-40
Insect cell nuclei soak buffer	50 mM Tris-HCl pH 7.4 400 mM KCl 10% glycerol
Kanamycin stock solution (25 mg/ml)	0.25 g ampicilin water to 10 ml, filter sterilized

Phenol/CIA	50% CIA 50% phenol (TE(10, 0.1)-equilibrated)
Protein gel destain	7% (v/v) ethanol 5% (v/v) acetic acid
Protein gel fix	45% (v/v) ethanol 5% (v/v) acetic acid
Protein gel loading buffer (PGLB)	0.4 mg/ml bromophenol blue 125 mM BisTris-HCl pH 6.8 20% (v/v) glycerol 4% (w/v) SDS 10% (v/v) β -mercaptoethanol
Protein gel running buffer (PGRB)	50 mM Tris (pH not adjusted) 0.38 M glycine 0.1% (w/v) SDS
Protein gel stain	5 mg/ml Coomassie Blue R in Protein gel fix
Sequencing buffer	80 mM Tris-HCl pH 8.8 2 mM $MgCl_2$
T4 DNA ligation buffer	50 mM Tris-HCl pH 7.5 10 mM $MgCl_2$ 1 mM DTT 1 mM ATP 0.025 mg/ml BSA
T4 PNK reaction buffer	70 mM Tris-HCl pH 7.6 10 mM $MgCl_2$ 5 mM DTT
TE buffer (10, 1.0)	10 mM Tris-HCl pH 8.0 1 mM EDTA-NaOH pH 8.0
TE buffer (10, 0.1)	10 mM Tris-HCl pH 8.0 0.1 mM EDTA-NaOH pH 8.0
Tetracycline stock solution (10 mg/ml)	0.1 g tetracycline water to 10 ml, filter sterilized
TRR mix	Applied Biosystems (Contains dye terminators; dNTPs; AmpliTaq DNA polymerase, FS; rTth pyrophosphatase; $MgCl_2$; Tris-HCl pH 9.0 buffer)
Vent DNA polymerase reaction buffer (ThermoPol buffer, NEB, B90045)	20 mM Tris-HCl pH 8.8 10 mM KCl 10 mM $(NH_4)_2SO_4$ 2 mM $MgSO_4$ 0.1% Triton X-100
Western blot buffer	25 mM Tris (pH not adjusted) 192 mM glycine

2.1.3 Enzymes

Restriction endonucleases:

Enzyme	Manufacturer	Number
<i>Afl</i> II	New England Biolabs	R0520S
<i>Aha</i> NI	New England Biolabs	R0514S
<i>Apa</i> I	New England Biolabs	R0114S
<i>Avr</i> II	New England Biolabs	R0174S
<i>Bam</i> HI	New England Biolabs	R0136L
<i>Bcl</i> I	New England Biolabs	R0160S
<i>Bgl</i> II	New England Biolabs	R0143L
<i>Bss</i> H II	New England Biolabs	R0199S
<i>Bst</i> E II	New England Biolabs	R0162S
<i>Bst</i> Z17 I	New England Biolabs	R0594L
<i>Cla</i> I	New England Biolabs	R0197S
<i>Dpn</i> I	New England Biolabs	R0176L
<i>Dra</i> I	New England Biolabs	R0129L
<i>Eag</i> I	New England Biolabs	R0129S
<i>Eco</i> R I	New England Biolabs	R0101L
<i>Eco</i> R V	New England Biolabs	R0195S
<i>Fsp</i> I	New England Biolabs	R0135S

<i>Hind</i> III	New England Biolabs	R0104L
<i>Kpn</i> I	New England Biolabs	R0142L
<i>Mfe</i> I	New England Biolabs	R0589S
<i>Nar</i> I	New England Biolabs	R0191S
<i>Nco</i> I	New England Biolabs	R0193L
<i>Nde</i> I	Fermentas	ER0581
<i>Nhe</i> I	New England Biolabs	R0131S
<i>Not</i> I	New England Biolabs	R0189L
<i>Nsi</i> I	New England Biolabs	R0127S
<i>Pae</i> I	Fermentas	ER0601
<i>Pme</i> I	New England Biolabs	R0560S
<i>Pst</i> I	New England Biolabs	R0140L
<i>Pvu</i> II	New England Biolabs	R0151L
<i>Rsr</i> II	New England Biolabs	R0501S
<i>Sma</i> I	New England Biolabs	R0141S
<i>Sna</i> B I	New England Biolabs	R0130S
<i>Spe</i> I	New England Biolabs	R0133S
<i>Sph</i> I	New England Biolabs	R0182S
<i>Ssp</i> I	Stratagene	501070
<i>Stu</i> I	New England Biolabs	R0187S
<i>Xba</i> I	Fermentas	ER0681
<i>Xho</i> I	Fermentas	ER0691
<i>Xma</i> I	New England Biolabs	R0180S

DNA modifying enzymes and polymerases:

Enzyme	Manufacturer	Number
Alkaline phosphatase from calf intestine (CIP)	Roche	713 023
DNase I	Invitrogen	18047-019
Micrococcal nuclease (MNase)	Amersham	E70196Y
<i>Pfu</i> DNA polymerase	Promega	M7741
T4 DNA ligase	New England Biolabs	M0202L
T4 polynucleotide kinase (PNK)	New England Biolabs	M0201L
Taq DNA polymerase	New England Biolabs	M0273S
Vent DNA polymerase	New England Biolabs	M0254L

Other enzymes:

Enzyme	Manufacturer	Number
Chymotrypsin A4 (bovine pancreas)	Boehringer Mannheim	103314
Elastase	Boehringer Mannheim	1027891
Enterokinase, light chain	New England Biolabs	P8070S
Ribonuclease A, RNase A	Sigma	R5500
Subtilisin	Fluka	85968
TEV protease	self-made	
Trypsin (bovine pancreas)	Boehringer Mannheim	109819

2.1.4 Bacterial strains and media

In the course of this work, the following cloning and expression *E. coli* strains were used:

Bacterial strain	Genotype	Resistance	Description
TOP10	F' { <i>proAB</i> , <i>lacIq</i> , <i>lacZΔM15</i> , Tn10(Tet ^R)} <i>mcrA</i> , Δ(<i>mrr hsdRMS mcrBC</i>), <i>P80lacZΔM15</i> , Δ <i>lacX74</i> , <i>deoR</i> , <i>recA1</i> , λ- <i>araD139</i> , Δ(<i>ara-leu</i>)7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str ^R), <i>endA1</i> , <i>nupG</i>	streptomycin	General purpose cloning strain.

BW23473 (8B)	<i>ΔlacI69 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMtuI)::pir⁺</i>		Strain used for the low copy number (~15 per cell) replication of pUNI cloning vectors with the conditional origin of replication (requires <i>pir</i> gene).
DH10MultiBac	<i>F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galKλ⁻ rpsL nupG /pMON14272 / pMON7124 v-cath⁻ chiA⁻</i>		Strain containing baculovirus DNA (bacmid), used for recombinant baculovirus production by site-specific transposition. It is based on DH10Bac™ <i>E. coli</i> and has a deletion of <i>v-cath</i> and <i>chiA</i> genes removing a native baculovirus protease (Berger et al, 2004)
BL21(DE3)	<i>F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm (DE3)</i>		Expression strain, B-type restriction and methylation abolished. Deficient in Lon and OmpT proteases. DE3 lysogen, contains a gene for T7 RNA polymerase (produced upon IPTG induction).
BL21(DE3) pLysS	<i>F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm (DE3) pLysS(Cm^R)</i>	chloramphenicol	BL21(DE3) derivative, contains pLysS plasmid (with <i>p15A</i> origin, compatible with plasmids containing the <i>ColE1</i> or <i>pMB1</i> origin like <i>pUC</i> - or <i>pBR322</i> -derived plasmids) that constitutively expresses T7 lysozyme gene from a cat-promoter (to reduce basal level expression of the gene of interest).
BL21 Star pLysS	<i>F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm rne131 (DE3) pLysS(Cm^R)</i>	chloramphenicol	BL21(DE3) derivative, contains pLysS plasmid, deficient in RNase E (decreased degradation of mRNA).
Rosetta pLysS	<i>F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm lacY1 pLysSRARE (<i>argU, argW, ileX, glyT, leuW, proL</i>) (Cm^R)</i>	chloramphenicol	BL21 lacZY (Tuner™) derivative designed to enhance the expression of eukaryotic proteins that contain codons rarely used in <i>E. coli</i> . Supplies tRNAs for the codons AUA, AGG, AGA, CUA, CCC, GGA on a pLysS plasmid. The tRNA genes are driven by their native promoters.

The following media were sterilized in an autoclave for 20 min at 121° C. Antibiotics were added as required at the following concentrations: ampicillin 100 µg/ml, chloramphenicol 34 µg/ml, kanamycin 50 µg/ml, tetracyclin 12 µg/ml, gentamycine 7 µg/ml, and streptomycine 20 µg/ml.

Growth medium:	Composition:
2xTY	1.6% (w/v) bacto tryptone 1.0% (w/v) yeast extract 0.5% (w/v) NaCl
TYE-agar plates	1.6% (w/v) bacto tryptone 1.0% (w/v) yeast extract 0.5% (w/v) NaCl 1.5% (w/v) agar
SPM	1.6% (w/v) bacto tryptone 1.0% (w/v) yeast extract 0.5% (w/v) NaCl 1M sorbitol 2.5 mM betaine

2.1.5 Insect cell culture material and equipment

The insect cell line used in this work is **Sf21**, cloned from pupal ovarian tissue of the Fall Armyworm *Spodoptera frugiperda* by Vaughn et al, 1977.

All insect cell culture handling was carried out in the insect cell culture room, thermostated at 27° C, using the following equipment:

Equipment	Manufacturer	Use
Ultrasafe 48	Faster	sterile hood
Pipetus Akku	Hirschmann	liquid handling
Rotina 35	Hettich	centrifuge
20P, 200P	Art	sterile tips
SF-900 II Medium with L-Glutamine	Gibco	cell culture media
Spinner Flasks	Wheaton	cell growth
CellFECTIN® Reagent	Invitrogen	initial transfection
Agar Plaque Plus™ Agarose	PharMingen	plaque assays
Incubator	Memmert	incubation
Shaking incubator	New Brunswick Scientific	cell growth
Gefran 1000 controller	Gefran-Acone	temperature control
Axiovert 25	Zeiss	microscope
Filtron Ultrapump II	Masterflex	media pumping
Biostat B	B.Braun Biotech	bioreactor

2.1.6 Cloning and expression vectors

Cloning vectors:

Vector	Description
pIBHO	Small medium to low copy number plasmid carrying a kanamycin-resistance gene and containing a multiple cloning site designed for manipulation of large human TAF genes.
pUNI10	Small plasmid with a conditional (R6Kγ) origin of replication that replicates only in 8B <i>E. coli</i> strain (expresses the <i>pir</i> ⁺ gene, resulting in low copy number of pUNI10). It is used for manipulation of human TAF1 gene (and other large constructs) and has a kanamycin-resistance gene.

pCR2.1-TOPO	Vector for PCR product subcloning, a part of the TOPO-TA cloning kit (Invitrogen). It is supplied linearized with single T overhangs which have a topoisomerase I covalently attached. It has an ampicillin and kanamycin resistance genes.
-------------	---

Bacterial expression vectors:

Vector	Description
pET3a	Component of the bacterial PET expression system carrying an ampicillin-resistance gene. It has a T7 phage promoter, which is recognized only by the T7 RNA polymerase (encoded in <i>E. coli</i> BL21(DE3) genome) but not by the host RNA polymerase (Studier et al., 1990).
pET15b	Plasmid based on pET3, with an N-terminal His-tag coding sequence, carrying an ampicillin-resistance gene.
pET28a	Plasmid based on pET3, with an N- and C-terminal His-tag and T7-tag coding sequence, carrying a kanamycin-resistance gene.

Insect cell expression vectors:

Vector	Description
pFastBac™ HT	Component of the Bac-toBac® baculovirus expression system (Invitrogen). Contains polyhedrin promoter, a multiple cloning site, ampicillin and gentamycin resistance genes and Tn7 recombination sites.
pFBDiesel (pDiFB)	Dual expression cassette plasmid with two multiple cloning sites based on pFastBac™ Dual (Invitrogen), designed by Dr. Imre Berger for easy cloning of multiple expression cassettes (using <i>Sna</i> B I and <i>Avr</i> II sites).
pFBDO	Dual expression cassette plasmid with two multiple cloning sites based on pFBDiesel. One site has a polyhedrin promoter and the second one p10 promoter. Designed by Dr. Imre Berger to use for human TAF genes with <i>Rsr</i> II and <i>Bst</i> E II cloning sites.

2.1.7 Oligonucleotides

Sequencing primers used in this work:

Primer	Sequence	Description
STO313	CCCTCTAGAAATAATTTTGTTTAACTTT	forward primer for PCR screening and sequencing of pET-based plasmids
STO314	GCTTCCCAGTCAGTGCTCC	reverse primer for PCR screening and sequencing of pET-based plasmids
BDO21	GCATCTGTGCGGTATTTACACC	forward sequencing primer for pIBH1, pIBH2, pIBHO plasmids
BDO22	GCAGTTTCATTTGATGCTCGATGAG	reverse sequencing primer for pIBH1, pIBH2, pIBHO plasmids
polf	TTTACTGTTTTCGTAACAGTTTTG	forward sequencing primer for pFastBac-based plasmids (polyhedrin promoter MCS)
polr	GGTTTTTAAAGCAAGTAAACCTC	reverse sequencing primer for pFastBac-based plasmids (polyhedrin promoter MCS)
hTAF250A	GGGACCCGGCTGCGATTTGCTGCTGC	internal, forward sequencing primer for human TAF1 gene
hTAF250B	GGGGAGCTTGCAGCCTTGC	internal, forward sequencing primer for human TAF1 gene
hTAF250C	AGAAAGAGGAAGAAGAAGCACCG	internal, forward sequencing primer for human TAF1 gene
hTAF250D	GGCACTGATCTTCTGGCTGATG	internal, forward sequencing primer for human TAF1 gene
hTAF250E	TGACTAGGAATGCGATGGCTTAC	internal, forward sequencing primer for human TAF1 gene
hTAF250F	TCC TGC TGT GGA ATT ACG GC	internal, forward sequencing primer for human TAF1 gene
hTAF250G	GCCATACATCTCCTTTCCTGGG	internal, forward sequencing primer for human TAF1 gene
hTAF250H	CTCTGCGCTGACTTCAAACG	internal, forward sequencing primer for human TAF1 gene
hTAF250I	AACAGATGCAGACCTTCGTCC	internal, forward sequencing primer for human TAF1 gene
hTAF250K	GCGGAAGGAACACTACAGCGAA	internal, forward sequencing primer for human TAF1 gene

hTAF250L	GAAACAGCTAATTGAGAGTGCGG	internal, forward sequencing primer for human TAF1 gene
hTAF250M	AGTGCAACCTACAATGGGCC	internal, forward sequencing primer for human TAF1 gene
hTAF250N	GCCCAGGAGATTGTGAACGTCTG	internal, forward sequencing primer for human TAF1 gene
hTAF250O	CTTATGTCTGAAGGAGAAGATGATG	internal, forward sequencing primer for human TAF1 gene
hTAF250P	GGGCCGAGCGTACTAAGCCAGGTCC	internal, forward sequencing primer for human TAF1 gene
hTAF250Bnew	GATGAAGATGACTATGATG	internal, forward sequencing primer for human TAF1 gene
hTAF250Cnew	GGAGTGCTCAGTAGAATCAG	internal, forward sequencing primer for human TAF1 gene
hTAF250Dnew	GCATTGGGAGGATGATATC	internal, forward sequencing primer for human TAF1 gene
hTAF250Erev	CTCTCTTGTCTCTCATCTTGG	internal, reverse sequencing primer for human TAF1 gene (for a region difficult to sequence in the forward direction)
MM250M	GCATCTCGAGACATACTGAGGGATG	internal, forward sequencing primer for human TAF1 gene
MM150Ar	GTTGTGGGATTTGTGGAAGCTG	internal, forward sequencing primer for human TAF2 gene
MM150B	GGGAAGTATGGCAGAGAGAG	internal, forward sequencing primer for human TAF2 gene
MM150C	GATTATAGATGAGACACCTTGT	internal, forward sequencing primer for human TAF2 gene
MM150Dr	GTGGTTTCCACATCTGGG	internal, forward sequencing primer for human TAF2 gene
MM150E	GGAAAAATTCCTACTCCAG	internal, forward sequencing primer for human TAF2 gene
MM150F	GAATATGGAAAACTTCTCCG	internal, forward sequencing primer for human TAF2 gene
MM150G	GAGTCAGTAGCAGGCAACCAAG	internal, forward sequencing primer for human TAF2 gene
MM150FG	GAACATGGAGTCTCCCTTATG	internal, forward sequencing primer for human TAF2 gene
MM150EF	GAAGACTATGCCAGTTGCAATG	internal, forward sequencing primer for human TAF2 gene
MM150Frev	GAGACC TTAATA CCATCC TAAG	internal, reverse sequencing primer for human TAF2 gene (for a region difficult to sequence in the forward direction)
TBPout3'	AATCTACAGAATGATCAAACCC	inside-out sequencing primer for human TBP gene (3'-region)
TBPcout5'	CGCTTGGGATTATATTCGGCG	inside-out sequencing primer for human TBPc gene (5'-region)
250out3'	CGAGGAGGACAGTGAGGATTTC	inside-out sequencing primer for human TAF1 gene (3'-region)
250out5'	CCTCCAGCGGAATCTTCGTCGC	inside-out sequencing primer for human TAF1 gene (5'-region)
150out3'	GCACAAACACAAGCATAAGCATGAC	inside-out sequencing primer for human TAF2 gene (3'-region)
150out5'	TATGTTGTTGATGCAGACGACCTG	inside-out sequencing primer for human TAF2 gene (5'-region)

PCR primers used in this work:

Primer	Sequence	Description
MM001	GGAATTCATATGGGTAACCACAACA GGAAGAAAGGAGACAAGGGCTTTGA AAGCCC	forward primer for construction of pIBHO150
MM002	CGGGATCCCGGACCGGTCTGAAAGGG AAGGAGAACG	reverse primer for construction of pIBHO150
MM003	CGGGATCCCGGACCGTTCATCAGAGT CCAAGTCACTG	reverse primer for construction of pUNI10h250Rsr
MM004	ATGGACCTGAGAGTCAGTATACTAAG ACTGCCAG	forward primer for construction of pUNI10h250Rsr
MM005	GTGCGGCATGCGCGCCTTGAGCCTGG CGAACAGTTCCG	forward primer for construction of pUNI10BstCBPSTREP250Rsr
MM006	GCCCCGAGGATGAGATTTTCTTAAA GCCGTTGGCTGCTGAGACGGCTATGA AATCTTTTTCCATCGTCGCTTGTCAT GGTACCTTCTCGAGCCCGGGAATT CCAGATAACTTCG	gene SOEing primer (part 1) for introduction of CBP-STREP tag to the N-terminus of TAF1 (construction of pUNI10BstCBPSTREP250Rsr)
MM007	GGAATTCATATGGGAGCCCTTGTGCG TCGTCGTCGTCGGCGCCTTTTTTGAAC TGGGGTGGCTCCAGTAGCTGCCCC GGAGGATGAGATTTTCTTAAAGCGGT TGG	gene SOEing primer (part 2) for introduction of CBP-STREP tag to the N-terminus of TAF1 (construction of pUNI10BstCBPSTREP250Rsr)
MM2501a	GCATCGATCGGACCGTCATTAACAG G GTTGAAGGTGGCGC	reverse primer for the construction of pFBDO250cons1

MM2501b	GACAAGGGCTCCCATATGGGACCCG	forward primer for the construction of pFBDO250cons1 and pUNI10250cons1
MM2502a	GATCGATCCATATGACAGAAGAACAG GAGGAGGAGTTGG	forward primer for the construction of pFBDO250cons2 and pUNI10250cons1
MM2502b	GCATGCATCGGACCGTCATTATTCATC AGAGTCCAAGTCACTG	reverse primer for the construction of pFBDO250cons2
MM2503a	GATCGATCCATATGCAGCCAGGTCCC CACTCAG	forward primer for the construction of pFBDO250cons3 and pUNI10250cons1
MM2503b	GATCGATCCGGACCGTCATTAGGTCA TTGGGTCCAGG	reverse primer for the construction of pFBDO250cons3
MM2501aNEW	GCATCGATGGATCCTCATTAAACAGG GTTGGAAGGTGGCGC	reverse primer for the construction of pUNI10250cons1
MM2502bNEW	GCATGCATGGATCCTCATTATTCATCA GAGTCCAAGTCACTG	reverse primer for the construction of pUNI10250cons2
MM2503bNEW	GATCGATCGGATCCTCATTAGGTCA GGGTCCAGG	reverse primer for the construction of pUNI10250cons3
MMKan5	GGAATTCCATATGGGTAACCGGTAAG ACACGACTTATCG	forward primer for the amplification of kanamycine resistance gene in pIBHO
MMKan3	GGAATTCCATATGGGTTACCAAACGT ACTGCGGTAAG	reverse primer for the amplification of kanamycine resistance gene in pIBHO
MutD1	CATACACTGCAAATTGAAGAAGAAAA TCTATACTCCAGGGATCCAACAGCCT TAAACATGATATACCC	forward mutagenesis primer for construction of TAF2 TEV site mutant D version 1
MutD2	GGGTATATCATGTTTAAGGCTGTTGG ATCCCTGGAAGTATAGATTTTCTTCTT CAATTTGCAGTGTATG	reverse mutagenesis primer for construction of TAF2 TEV site mutant D version 1
MutD1new	CCCTGCCATTCCAAAAGTAGAAGGGA AAATCTATACTCCAGGGATCCAATA AAAAGAAAAAATCCCACTG	forward mutagenesis primer for construction of TAF2 TEV site mutant D version 2
MutD2new	CAGTGGGATTTTTTCTTTTTATTGGA TCCCTGGAAGTATAGATTTTCCCTTCT ACTTTTGGAAATGGCAGGG	reverse mutagenesis primer for construction of TAF2 TEV site mutant D version 2
TAF2An	GATCGACCATGGGAGACAAGGGCTTT G	forward PCR primer for construction of bacterial expression vector for TAF2 mutant A and TAF2 TR-4
TAF2AcBact	GAGATCCTCGAGTTTCTCCTTAGATT AAG	reverse PCR primer for construction of bacterial expression vector for TAF2 mutant A and TAF2 TR-5
TAF2Tr4c	GAGATCCTCGAGCCTTCTACTTTTGG ATG	reverse PCR primer for construction of bacterial expression vector for TAF2 TR-4
TAF2Tr5n	GATCGACCATGGGAAATAAAAAAGAA AAAAATCCCACTG	forward PCR primer for construction of bacterial expression vector for TAF2 TR-5
MM150repair	GCAACCATTGCAGCATCTACTG	reverse PCR primer for repair of the non-silent mutation in TAF2 gene from pIBHO150
MutA1	TGGTTCTTAATCTAAAGGAGAAAGAA AATCTATACTCCAGGGATCCAAAGC TGTCTTGAATCCTACC	forward mutagenesis primer for construction of TAF2 TEV site mutant A
MutA2	GGTAGGATTCAAGACAGCTTTGGATC CCTGGAAGTATAGATTTTCTTCTCCT TTAGATTAAGAACCA	reverse mutagenesis primer for construction of TAF2 TEV site mutant A
MutB1	AGTCAGTGTGAATAATGAAGTTAGAG AAAATCTATACTCCAGGGATCCACTT TGGATAACTTAAATCCTGATG	forward mutagenesis primer for construction of TAF2 TEV site mutant B
MutB2	CATCAGGATTTAAGTTATCCAAAGTG GATCCCTGGAAGTATAGATTTTCTCTA ACTTCATTATTCACACTGACT	reverse mutagenesis primer for construction of TAF2 TEV site mutant B
TAF2tailN	GATCGACCATGGGAAAAGCTGTCTTG G	forward PCR primer for construction of pET28aTAF2tail
TAF2tailC	GATCGACTCGAGTCAGTCTGAAAGGG	reverse PCR primer for construction of pET28aTAF2tail

Other oligonucleotides used in this work:

Name	Sequence	Description
MartinDfwB	ACCCCCGAGCCAAGGGGGA	5'-biotinilated forward oligonucleotide for TAF2 BiaCore experiments, forward
MartinDrevB	TCCCCCTTGGCTCGGGGGT	5'-biotinilated forward oligonucleotide for TAF2 BiaCore experiments, reverse
MartinNEGfwB	ACCCGCTTAGCAGGGGGA	5'-biotinilated forward oligonucleotide for TAF2 BiaCore experiments, forward (negative control)
MartinNEGrevB	TCCCCCTGCTAAGCGGGT	5'-biotinilated forward oligonucleotide for TAF2 BiaCore experiments, reverse (negative control)
AdMLPfw	TGACCGGGTGTTCCTGAAGGG GGGCTATAAAAGGGGGTGGGG GCGCGTTCGTCCTCACTC	band-shift oligonucleotide based on AdML promoter, forward
AdMLPrev	GAGTGAGGACGAACGCGCCC CCACCCCTTTTATAGCCCCC TTCAGGAACACCCGGTCA	band-shift oligonucleotide based on AdML promoter, reverse
MartinDfw	ACCCCCGAGCCAAGGGGGA	band-shift oligonucleotide based on Martin et al, 1999, fig 4D, oligonucleotide "d", forward
MartinDrev	TCCCCCTTGGCTCGGGGGT	band-shift oligonucleotide based on Martin et al, 1999, fig 4D, oligonucleotide "d", reverse
MartinNEGfw	ACCCGCTTAGCAGGGGGA	band-shift oligonucleotide for negative control of TAF2 band-shifts, forward
MartinNEGrev	TCCCCCTGCTAAGCGGGT	band-shift oligonucleotide for negative control of TAF2 band-shifts, reverse
cyclin01	AGGGAGCAGTGCGGGGTTTAA ATCTGAGGCTAGGCTGGCTCTT CTCGGGC	band-shift oligonucleotide based on Cyclin B1 core promoter, forward
cyclin02	CGCCGAGAAGAGCCAGCCTA GCCTCAGATTTAAACCCCGCA CTGCTCCCT	band-shift oligonucleotide based on Cyclin B1 core promoter, reverse
core01	GATCGAGGGCGCCCTATAAAG ACAGTTTGCAGAGCAGTGCCC TCATTCTCGATCGATCGATCGA ACGGAACGGACGTGGATCGAT CGA	band-shift oligonucleotide based on a synthetic core promoter sequence designed during this work, forward
core02	TCGATCGATCCACGTCCGTTCC GTTTCGATCGATCGATCGAGAA TGAGGGCACTGCTCGAAAC TGTCTTTATAGGGCGCCCTCGA TC	band-shift oligonucleotide based on a synthetic core promoter sequence designed during this work, reverse
Kauff01	ACATCAGAGCCCTCATTCTGG AGA	band-shift oligonucleotide based on the sequence used for affinity purification of TAF2 from HeLa cells (Kaufmann et al, 1997), forward
Kauff02	TCTCCAGAATGAGGGCTCTGA TGT	band-shift oligonucleotide based on the sequence used for affinity purification of TAF2 from HeLa cells (Kaufmann et al, 1997), reverse
GAG01	GTACCGAGCTCGGATCCAGG AACTCGAGCATGCATCTAG	band-shift containing "GAG" core sequence (used in Martin et al, 1999), forward
GAG02	CTAGATGCATGCTCGAGTTCTC TGGGATCCGAGCTCGGTAC	band-shift containing "GAG" core sequence (used in Martin et al, 1999), reverse

2.1.8 DNA and protein markers

DNA markers:

λ HindIII DNA marker: 23 130 bp 9 416 bp 6 557 bp 4 361 bp 2 322 bp 2 027 bp
λ HindIII-EcoRI DNA marker: 21 226 bp 5 148 bp 4 973 bp 4 268 bp 3 530 bp 2 027 bp 1 904 bp 1 584 bp 1 375 bp 947 bp 831 bp 564 bp 125 bp
pST14 MspI-EcoRV DNA marker: 540 bp 489 bp 404 bp 320 bp 242 bp 190 bp 147 bp 110 bp 95 bp 67 bp 34 bp
2-Log DNA ladder (New England Biolabs, Cat. No. N3200S): 10 kbp 8 kbp 6 kbp 5 kbp 4 kbp 3 kbp 2 kbp 1500 bp 1200 bp 1000 bp 900 bp 800 bp 700 bp 600 bp 500 bp 400 bp 300 bp 200 bp 100 bp
1 kb DNA ladder (New England Biolabs, Cat. No. N3232S): 10 kbp 8 kbp 6 kbp 5 kbp 4 kbp 3 kbp 2 kbp 1.5 kbp 1 kbp 0.5 kbp
100 kb DNA ladder (New England Biolabs, Cat. No. N3231S): 1 517 bp 1 200 bp 1000 bp 900 bp 800 bp 700 bp 600 bp 500/517 bp 400 bp 300 bp 200 bp 100 bp

Protein markers:

Unstained Protein Molecular Weight Marker (Fermentas, Cat. No. SM0431): 116.0 kDa 66.2 kDa 45.0 kDa 35.0 kDa 25.0 kDa 18.4 kDa 14.4 kDa
Prestained Protein Molecular Weight Marker (Fermentas, Cat. No. SM0441): 120 kDa 86 kDa 47 kDa 34 kDa 26 kDa 20 kDa
Broad Range molecular weight protein marker (BR, BioRad, Cat. No. 72807A): 200 kDa 116 kDa 97.4 kDa 66 kDa 45 kDa 31 kDa 21.5 kDa 14.5 kDa 6.5 kDa

2.1.9 Commercial kits

Name	Supplier	Catalog no.	Usage
QIAprep [®] Spin Miniprep Kit	Qiagen	27104	small-scale plasmid purification
E.Z.N.A. [®] Plasmid Miniprep Kit I	Peqlab	12-6942-02	small-scale plasmid purification
QIAGEN [®] Plasmid Midi Kit	Qiagen	12143	middle-scale plasmid purification
QIAquick [®] Gel Extraction Kit	Qiagen	28704	extraction of DNA fragments from agarose gels, purification of PCR products
TOPO TA Cloning [®] Kit	Invitrogen	K4500-01	subcloning of PCR products

2.1.10 Equipment

General apparatus:

Equipment	Supplier	Usage
ABI PRISM 310 DNA sequenator	Applied Biosystems	DNA sequencing
Balances	Mettler	Weight measuring
BIORAD Gene Pulser® II	BioRad	Electroporation
CD Spectropolarimeter J710	JASCO	Circular dichroism spectrometry
Centrifugal filter devices, Centricon	Amicon (Millipore)	Protein concentration
Centrifuge tubes, plastic	Sorvall	Separation
Centrifuge tubes, Corex glass	Corning	Separation
Deep freezer (-80° C)	REVCO	Storage of cells, proteins and DNA
Dialysis tubing	Spectrum, Roth	Dialysis
Electrophoresis gel boxes	BioRad	Electrophoresis
Electrophoresis loading syringes	Hamilton	Electrophoresis
Electrophoresis power supply 2197	LKB Bromma	Electrophoresis
EmulsiFlex-C5 (cell cracker)	Avestin	Mechanical cell lysis
Filters, sterilizing 22, 45 µm	Sartorius	Sterile Filtration
Freezer (-20° C)	Burkhalter	Storage of proteins and DNA
Heating block Dri-Block DB-3	Techne	Enzyme incubation, inactivation
UV-transilluminator	Hilpert electronics	UV-illumination
Incubator B5050E	Heraeus	Incubation of bacteria agar plates
Incubator-shaker G25	New Brunswick scientific	Incubation of bacteria liquid cultures
JASCO CD spectrometer	J715	JASCO Corp.
MilliQ-water	Millipore	Water Purification System
Peltier Thermal Cycler PTC-200	MJ Research	PCR
pH Meter PH M 85 Precision	Radiometer Copenhagen	pH measurements
Shaker Certomat WR	Bender & Hobein	Coomassie staining/ destaining
Ultrasonic processor XL	Heart Systems	Homogenizing cell lysates
Spectrophotometer Cary 1E	Varian	UV-absorption spectroscopy
Spectrophotometer Novaspec II	Parmacia	OD of bacteria liquid cultures measurement
Spiramax	Denley	Mixing
UV Transilluminator 211 Macrovue	LKB Bromma	Ethidium bromide visualization
Vortex Genie 2TM	Bender & Hobein	Mixing
Waterbath Tempette TE-8D	Techne	Coomassie staining / destaining
Waterbath Haake K20	Haake	Ligation (16° C)

Centrifugation processes are documented using the formalism: (rpm, t, T, c, r). This stands for a centrifugation at rotations per minute 'rpm', for the time 't', at the temperature 'T', in the centrifuge 'c' using the rotor 'r'. An example is (12000, 20 min, 4° C, RC 26, SS34).

Centrifuge	Rotor / Optical system	max. radius
Centrifuge 5417, Eppendorf ("benchtop")	fixed angel	7.9
Minifuge T, Heraeus ("tabletop")	swinging bucket	19.2
Sorvall RC 26 plus, Sorvall ("RC 26")	SS34 (HFA 2250, Heraeus, fixed 34° angle)	10.9
	GSA (HFA 14290, Heraeus, fixed 25° angle)	14.1
	SLA3000 (IIFA 12500, Heraeus, fixed 20° angle)	16.2
Beckman XL-I analytical ultracentrifuge	An50 Ti / Optima™ XL-I scanning UV/VIS absorbance optical system	7.2

FPLC chromatography was carried out using an ÄKTA FPLC system (Amersham-Pharmacia) or conventional chromatography Gilson equipment:

Equipment	Model	Supplier
ÄKTA™	FPLC	Amersham-Pharmacia
Dual wavelength UV detector	112	Gilson
Fraction collector	202	Gilson
Peristaltic pump	Minipuls 3	Gilson
Recorder		Kipp-Zonen
Fittings, Tubings		Omnifit
Empty columns	XK series	Amersham-Pharmacia
Empty columns	Econo-Pac	Biorad
HPLC-Pump	305/306	Gilson
Mixing Chamber	811	Gilson
Manometric module	802	Gilson
Dual wavelength UV detector	119	Gilson
Fraction collector	203	Gilson
Recorder		Kipp-Zonen
Superloop 50ml	M6	Amersham-Pharmacia
Fittings, Tubings		Rheodyne
Backwash, peristaltic pump	Minipuls 3	Gilson
Computer	GX50	DELL
Software, control	UniPoint Version 3.3	Gilson
HPLC system	BioSys 2000	Beckman

Chromatography resins used in this work:

Resin	Manufacturer	Number
TALON™ Sepharose CL-6B	Clontech	8901-3
Heparin-Sepharose CL-6B	Amersham-Pharmacia	17-0467-01
Hydroxyapatite	BioRad	157-0040
SP-Sepharose fast flow	Amersham-Pharmacia	17-0729-01
Superose 6	Amersham-Pharmacia	17-0489-01
Sephadex G-25 medium	Amersham-Pharmacia	17-0033-01
Q-Sepharose fast flow	Amersham-Pharmacia	17-0510-01
CM-Sepharose	Amersham-Pharmacia	17-0720-01
DEAE-Sephacel	Amersham-Pharmacia	17-0500-01

Chromatography columns used in this work:

Column	Size	Supplier
DEAE-5PW	21.5 mm x 150 mm	Tosohaas (TSK)
CM-3SW	21.5 mm x 150 mm	Tosohaas (TSK)
Mono-Q	5 mm x 50 mm	Amersham-Pharmacia
RESOURCE® S	6 ml	Amersham-Pharmacia
RESOURCE® S	1 ml	Amersham-Pharmacia
RESOURCE® Q	1 ml	Amersham-Pharmacia
RESOURCE® Phe	1 ml	Amersham-Pharmacia
Superdex 200 HR 10/30	10 mm x 300 mm	Amersham-Pharmacia

Crystallization equipment:

Equipment	Supplier	Use
Multiwell™ 24 well plates, Falcon Cat. No. 35-3226	Becton Dickinson	Crystallization
96 well plates, Greiner bio-one Cat. No. 609101	Huber + Co AG	Crystallization
Crystal clear tape Cat. No. XPX 2K	Henkel	Sealing
Microscope Typ 391354	Wild Heerbrugg	Visualization
Microscope, Typ MZ7 _s	Leica Microsystems	Visualization

Surface plasmon resonance (BiaCore) equipment:

Equipment	Supplier	Use
BiaCore™ 2000 (Functional Genomics Center Zürich)	BiaCore	SPR instrument
CM5 chip Cat. No. BR-1000-14	BiaCore	protein immobilization and SPR measurement
CM4 chip Cat. No. BR-1005-39	BiaCore	protein immobilization and SPR measurement
SA chip Cat. No. BR-1000-32	BiaCore	protein/DNA immobilization and SPR measurement
C1 chip Cat. No. BR-1005-40	BiaCore	protein immobilization and SPR measurement
Amine coupling kit Cat. No. BR-1000-50	BiaCore	covalent immobilization of ligands
Thiol coupling kit Cat. No. BR-1005-57	BiaCore	covalent immobilization of ligands

2.2 Recombinant DNA methods

2.2.1 Determination of DNA concentration

Concentrations of DNA samples were determined by UV-absorption spectroscopy. The absorbance (A) from 320 nm to 220 nm was measured against a corresponding buffer background sample. DNA concentrations (cDNA) were calculated with the formula $(A_{260}-A_{320}) = cDNA * \epsilon * d$ using the extinction coefficients $\epsilon = 0.03 \text{ cm}^{-1} \text{ mg}^{-1} \text{ ml}$ for short single stranded DNA (ssDNA, oligonucleotides) and $\epsilon = 0.02 \text{ cm}^{-1} \text{ mg}^{-1} \text{ ml}$ for double stranded DNA (dsDNA).

2.2.2 DNA polyacrylamide gel electrophoresis (DNA PAGE)

DNA fragments from 20 to 1000 bp were analyzed using a 10% (37.5:1 acrylamide:bisacrylamide), 1x TBE separating gel. The gel size was 10 cm x 10 cm x 0.5 mm (minigels). Samples were mixed with 0.2 volume of 6x DNA gel loading buffer. After loading the gels were run for 20 min at 10 W using 1x TBE as a running buffer. Afterwards gels were stained with ethidium bromide (0.5 $\mu\text{g}/\text{ml}$ in water). DNA bands were visualized under UV light (254nm).

2.2.3 Agarose gel electrophoresis

DNA fragments from 200 to 10 000 bp were analyzed using 0.8-1.5% agarose, 1 x TBE, separating gels. Samples were mixed with 0.2 volumes of 6x DNA gel loading buffer. After loading the gels were run for 30-60 min at 6-10 Volt/cm using 1x TBE as a running buffer. Afterwards gels were stained with ethidium bromide (0.5 $\mu\text{g}/\text{ml}$ in water). DNA bands were visualized under UV light (254nm).

2.2.4 Small scale alkaline lysis plasmid preparation (“miniprep”)

A method developed by Sambrook et al., 1989, and adapted in-house was used to assay transformed *E. coli* cells for the correct length of the vector insert. Up to 24 single colonies of *E. coli* cells freshly transformed with the appropriate plasmid were picked from a TYE agar plate

and used to inoculate 5 ml of 2xTY growth medium. The cultures were incubated overnight at 37° C in a shaking air incubator at 210 rpm. The overnight cultures were centrifuged (4000 rpm, 5 min, 22° C, tabletop) in 15 ml polypropylene tubes. The pelleted cells were resuspended in 100 µl lysis buffer and transferred to 1.5 ml Eppendorf tubes. 200 µl NaOH/SDS solution (0.2 M NaOH, 1% w/v SDS) was then added, and the tubes were shaken vigorously until they appeared to be translucent. After incubation at room temperature for 3 min, 300 µl of ice cold 5 M potassium acetate/2.5 M acetic acid solution were added to each sample and the tubes were shaken briefly. After centrifugation of the samples (13000 rpm, 5 min, RT, benchtop) the supernatant was extracted with 500 µl phenol/CIA (1:1). After ethanol precipitation, the DNA pellet was resuspended in 50 µl TE(10, 0.1) containing 40 µg/ml RNase A and incubated for 30 min at 37° C. The DNA was used for analytical scale restriction digestion.

Alternatively, small scale plasmid purification was done using the E.Z.N.A.[®] Plasmid Miniprep Kit I or QIAprep[®] Spin Miniprep Kit according to the manufacturer's instructions.

2.2.5 Medium scale alkaline lysis plasmid preparation ("midiprep")

For preparative DNA isolation, QIAGEN[®] Plasmid Midi Kit was used according to the manufacturer's instructions. The eluate was used for restriction digestion, sequencing, or cloning purposes.

2.2.6 Ethanol precipitation of DNA fragments

To concentrate them, DNA fragments were precipitated by adding 2.5 volumes of ice cold absolute ethanol and 0.1 volume of 3 M sodium acetate-acetic acid pH 5.2. After 5 min incubation on ice, the precipitated DNA was harvested by centrifugation (30', 13000 rpm, 22° C, benchtop). The DNA pellet was washed with 70% EtOH, air dried and resuspended in desired volume of TE (10, 0.1) or MiliQ water.

2.2.7 Spin column purification

Spin columns were constructed by placing siliconized glass wool in the base of a 1 ml plastic syringe, after which the syringe was filled with 500 μ l of Sephadex G25 medium, equilibrated in the appropriate buffer. The syringe was placed inside a 15 ml polypropylene tube, and pre-equilibrated by centrifugation (2000 rpm, 5 min, 22° C, tabletop). The sample (<150 μ l) was then applied to the top of the spun column. The syringe was placed into a 1.5 ml Eppendorf tube inside a polypropylene tube and sample was eluted by centrifugation as above.

2.2.8 Agarose gel purification

DNA fragments (< 2 pmol per 10 mm lane) in 1x DNA gel loading buffer were applied to 0.8-1.2% agarose minigels (8 cm x 8 cm x 0.5 cm, SeaChem ME grade, 1x TBE, 0.5 μ g/ml Ethidium bromide in the gel), and electrophoresed at 70 V for 30-60 min. The desired bands were excised while being visualized on a UV transilluminator. The agarose slices were then placed in a 1.5 ml safe lock Eppendorf tube and purification was continued using the QIAGEN Gel Extraction Kit according to the manufacturer's instructions.

2.2.9 Preparation of electrocompetent cells

Glycerol stocks of the *E. coli* strains were streaked on TYE agar plates. A single colony was used to inoculate 100 ml of the 2xTY medium. After overnight growth at 37° C and 250 rpm in a shaking incubator, 5 ml of this culture was used to inoculate 500 ml of 2xTY-medium. This 500 ml culture was grown at 37° C and 250 rpm until the optical density at 600 nm reached 0.4-0.6. The culture was then chilled on ice for 15-20 min, distributed into 10 ice-cold polypropylene tubes and centrifuged (4000 rpm, 10 min, 4° C, tabletop). The pelleted cells were gently resuspended in a total of 500 ml ice-cold sterile MilliQ water and centrifuged again. This washing step was repeated once, but in the volume of 250 ml. The cell pellets were then resuspended in a total of 10 ml ice-cold sterile 10% glycerol, pooled and centrifuged (4000 rpm, 10 min, 4° C, tabletop). The cell pellet was finally resuspended in 1 ml of ice-cold sterile 10% glycerol. Aliquots of 40 μ l were dispensed into sterile 1.5 ml Eppendorf tubes, flash frozen in

liquid nitrogen, and stored at -80°C for about 6 months. They yielded an efficiency of about 10^7 - 10^9 cfu/ μg of pUC19 plasmid.

2.2.10 Transformation of bacteria by heat shock

Plasmid DNA (0.1-2.0 μg , or 5-10 μl of the ligation reaction) was added to 200 μl of chemically competent cells, which had been thawed on ice. Cell suspension was incubated on ice for 30 min, then heat-shocked at 42°C for 30 seconds and immediately chilled on ice for 10-20 seconds. 500 μl of 2xTY media was added and the cells were incubated for 60 min at 37°C in a shaking air incubator at 210 rpm. 100 μl of the transformed cells were plated on TYE agar plates supplemented with the appropriate antibiotics and incubated for 12 to 16 h at 37°C .

2.2.11 Transformation of bacteria by electroporation

Plasmid DNA (5-50 ng), or ethanol-precipitated ligation reaction product resuspended in 10 μl of MilliQ water was added to 40 μl electrocompetent cells in an ice-cold sterile electroporation cuvette. The cell suspension was incubated on ice for 30-60 seconds. The electroporation apparatus (BIORAD Gene Pulser[®] II) was set as follows: 25 μF , 2.5 kV, and 200 Ω . The cells were electroporated by delivery of a pulse with the above settings and 1 ml of 2xTY medium was added. The cells were transferred to a 1.5 ml Eppendorf tube and incubated for 60 min at 37°C in a shaking incubator at 210 rpm. 100 μl of the transformed cells were plated onto TYE agar plates containing appropriate antibiotics and incubated for 12 to 16 hours at 37°C .

2.2.12 Restriction endonuclease digestion

For analytical purposes, 0.5 μg DNA or 1-5 μl of the plasmid “miniprep” were mixed with the appropriate enzyme reaction buffer (1x final concentration) and 0.5-1.0 U of the restriction endonuclease, in the final volume of 10 μl . In the case of double digests, care was taken not to exceed 10% glycerol concentration in the final reaction mixture, since it can lead to non-specific (“star”) activity. After 1-2 h incubation at 37°C , 2 μl of 6x DNA gel loading buffer

were added and the reaction products were analyzed by DNA PAGE or agarose gel electrophoresis, depending on the size of the expected DNA fragments.

For preparative purposes, 2-3 μg of DNA were digested with 10-20 units of restriction enzyme in a volume of 30 μl of the recommended New England Biolabs (NEB) buffer, which was supplemented with 10 mM DTT and 0.1 mg/ml BSA if required. After incubation at 37° C for 2 to 4 h, the digestion products were analyzed by DNA PAGE or agarose gel electrophoresis, depending on their size. Correctly digested products were mixed with the appropriate amount of 6x DNA gel loading buffer and electrophoresed on an agarose gel as described above for preparative gel purification.

2.2.13 DNA dephosphorylation

For dephosphorylation of digested plasmid vectors, calf intestinal alkaline phosphatase (CIP) was added to the heat inactivated digestion reaction. For vectors with protruding 5'-termini 0.1 unit of CIP and for blunt termini 1 unit of CIP was added to a 20 μl reaction mix. After 60 min incubation at 37° C the desired vector fragment was isolated by agarose gel purification or purified over a Qiagen DNA column. The DNA was used for ligation and subsequent transformation.

2.2.14 Ligation

For cohesive-ended ligation, 2 μl of 10x T4 DNA ligase buffer, 1 μl of 20 mM ATP, 2 μl of agarose gel purified vector DNA and 2 to 14 μl of agarose gel purified insert DNA (depending on insert size and concentration) were mixed in a volume of 20 μl . 4 units of T4 DNA ligase were then added. Reactions were incubated for 1-2 h at 22° C or overnight at 16° C and then used for plasmid transformation. For blunt-ended ligation, 40 units of T4 ligase were used.

2.2.15 DNA sequencing

Nucleotide sequencing of dsDNA was carried out by the dideoxy chain termination method (Sanger et al., 1977) using dideoxy nucleotide triphosphate terminators labeled base-specifically with different fluorescent dyes. 0.2 μg of plasmid DNA in 20 μl sequencing buffer

(80 mM Tris-HCl pH 8.8, 2 mM MgCl₂) together with 2 pmol of sequencing primer were PCR-amplified with dNTPs and labeled ddNTP mixture (TRR mix) as supplied by the manufacturer. Thermocycle Sequencing was performed for 25 cycles using the following PCR program:

96° C, 10''
50° C, 5''
60° C, 4'

The PCR mixtures were then ethanol precipitated by addition of 2 µl 3 M sodium acetate-acetic acid pH 5.2 and 50 µl absolute ethanol and centrifugation (13000 rpm, 30 min, 22° C, benchtop). The DNA pellet was washed once with 250 µl 70% EtOH, centrifuged (13000 rpm, 5 min, 22° C, benchtop) and finally air-dried. The dried pellet was resuspended in 25 µl MilliQ H₂O. The sample was analyzed by capillary electrophoresis onto an ABI PRISM 310 sequencing machine. The DNA sequence obtained was analyzed with the DNAMAN version 5.2.9 (Lynnon BioSoft) program.

2.2.16 Restriction fragment subcloning

For restriction fragment subcloning, both a plasmid containing the vector DNA and a plasmid containing the insert DNA to be subcloned were digested by the appropriate restriction enzymes as described. If possible, to facilitate vector purification and later colony screening, a plasmid that contained another insert that had a different size than the insert to be subcloned was chosen for the vector DNA. The restriction digests were analyzed by agarose gel electrophoresis. The digested plasmid containing the vector DNA was subsequently dephosphorylated to reduce the background, especially if it had blunt or identical “sticky” ends. Then, both vector DNA and insert DNA were purified followed by ligation and transformation. Colonies were screened afterwards by analytical restriction digestion of their plasmids purified by small scale alkaline lysis procedure. The sequence of the insert was confirmed by DNA sequencing.

2.2.17 PCR subcloning

PCR subcloning was used primarily if new sequences (for example, purification tags) or new restriction sites had to be added to the insert of interest. DNA for subcloning was amplified

by the polymerase chain reaction (PCR) (Saiki et al., 1988). In a reaction volume of 100 μ l, 10 to 100 ng of plasmid DNA containing the sequence to be amplified was incubated on ice in a 0.5 ml Eppendorf tube containing 0.5 μ M of each forward and reverse PCR primer, Vent DNA polymerase 1x reaction buffer and 0.25 mM dNTP. Vent DNA polymerase (1 unit) was added last and the samples were transferred to a thermocycler. The samples were incubated at 95° C for 2 min. PCR amplification was carried out using the following program: 5 cycles with 30 s at 95° C, 30 s at a temperature 5° C lower than the lowest melting temperature of the two primers, and at 72° C for 60 s per kilobase of expected PCR product followed by 20-25 cycles of 30 s at 95° C, 30 s at 50-60° C and 60 s per kilobase of expected PCR product at 72° C. Finally, the PCR reaction mixtures were incubated for 10 min at 72° C. DNA products were analyzed by either DNA PAGE or agarose gel electrophoresis, depending on the expected size of the PCR product. The reaction products were phenol/CIA extracted twice, CIA extracted and ethanol-precipitated, or alternatively, purified over a Qiagen DNA column. The DNA was resuspended in MilliQ water and used for restriction digestion and subsequent ligation.

2.2.18 Design of PCR primers

PCR primers were designed to have a melting temperature T_m of more than 50° C. Melting temperature (T_m) for the oligonucleotides was calculated using the following formula (Sambrook et al., 1989):

$$T_m = 81.5 - 16.6 * \log_{10}[\text{Na}^+] + 0.41(\%G+C) - 600/n$$

$[\text{Na}^+]$:	Concentration in mM
%G+C:	G/C-content in %
n:	Number of base pairs

Self-complementarity, two primer complementarity and mispriming were minimized with the program DNAMAN version 5.2.9 (Lynnon BioSoft). The PCR primers used in this study are listed in the section 2.1.7.

2.2.19 PCR subcloning using the TOPO TA cloning strategy

TOPO TA cloning strategy was used if conventional PCR subcloning was unsuccessful. The PCR product was prepared as described above, using the Vent DNA polymerase. Since the Vent polymerase doesn't leave single 3' A-overhangs necessary for the integration into the TOPO vector, they were produced by adding 1 U of the Taq DNA polymerase to the PCR product and incubating for 10 min at 72° C. The reaction was used immediately for the TOPO cloning, according to the manufacturer's instructions. Alternatively, PCR product was first purified by agarose gel electrophoresis and then 3' A-overhangs were produced using the Taq polymerase buffer supplemented with dATP (other conditions as above).

2.2.20 Gene SOEing method

Gene SOEing is the method for adding large DNA fragments (100-150 bp) to one end of a gene of interest. Since the extra DNA fragment is too large to be introduced using a single PCR primer, it is designed as two large primers with a 30% overlap. The PCR reaction is performed as described above, with the primer carrying the first half of the desired sequence and a conventionally designed primer for the other end. The resulting PCR product is purified using an agarose gel electrophoresis and used as a template for a second PCR reaction, with the primer carrying the second half of the desired sequence instead of the first one.

2.2.21 Quick change mutagenesis (adapted from S. Hahn lab)

A quick change mutagenesis procedure, originally designed by Stratagene and developed by Steven Hahn lab for the introduction of multiple mutations, was adapted for very large DNA insertions, necessary in the production of TAF2 TEV constructs described in the section 4.3.1. Two mutagenic primers with complementary sequences were designed using the software on Stratagene web pages (available at: <http://labtools.stratagene.com/Forms/SVLogin.php>). The software suggests the primer length based on calculated "energy cost" for generation of primer-template mismatches at specified extension temperature.

Two separate primer extension reactions were set up (one for each primer), containing:

water	41 μ l
10x Pfu reaction buffer	5 μ l
10 mM dNTP's	1 μ l
0.2 μ g/ μ l template DNA	1 μ l
10 μ M primer 1 or 2	1 μ l
2 u/ μ l Pfu polymerase	<u>1 μl</u>
	50 μ l

The following PCR program was used:

30 s 95° C → 5 x (30 s 95° C → 60 s 55° C → 2 min/kb 68° C) → 4° C

25 μ l from each of the two reactions were then combined and 1 μ l of the Pfu polymerase was added again. The mixture was incubated in the PCR using the same program, but for 18 cycles. 25 μ l of the second PCR reaction's product was taken and digested with 20 units of the methylation sensitive restriction enzyme *Dpn* I at 37° C for 5 h. 1 μ l of the final reaction was used for the transformation of the electrocompetent TOP10 *E. coli* cells. The transformed colonies were screened by restriction enzyme analysis (in the case when primers introduced an additional restriction site) and confirmed by DNA sequencing.

2.2.22 Annealing of the oligonucleotides

To anneal two complementary oligonucleotides, an equimolar concentration (in μ M range) was mixed together with 2 mM MgCl₂ in the volume of 50-200 μ l, followed by an incubation for 10 min at 90° C. The annealing took place during the cool-down phase in the thermoblock, until the temperature of the sample was equilibrated with the room temperature (3-4 hours).

2.2.23 DNA end labeling with ³²P

³²P-radioactive labeling reactions were performed with 50 pmol of the oligonucleotide in the reaction mixture containing 1x T4 polynucleotide kinase buffer, 5 mM DTT, 10 μ Ci of γ -³²P-ATP (10 mCi/ml) and 10 units of T4 polynucleotide kinase. The phosphorylation took place during incubation at 37° C for 15 min. The enzyme was subsequently removed by Phenol/CIA

and CIA extraction. The complementary oligonucleotides were annealed as described above, followed by spun column purification (G25M resin) for buffer exchange and small molecular weight contaminants removal.

2.3 Bacterial protein expression methods

For obtaining large protein quantities necessary for biochemical and structural studies, it is convenient to use *Escherichia coli* as an organism for overexpression. *E. coli* has an advantage of easy culture handling and a wide availability and versatility of cloning and expression vectors. Rapid culture growth and recovery of large amounts of the polypeptide product make *E. coli* expression systems the first choice in protein production.

The *E. coli* protein expression system used in this study is pET system from Novagen. It is based on the BL21(DE3) strain of *E. coli* which has a T7 virus RNA polymerase encoded in its genome under a IPTG-inducible (*lac*) promoter. The cells are transformed with the gene of interest in a vector from pET series. The gene is placed under a T7 promoter which makes expression possible to control.

2.3.1 Small scale expression

Before large scale expression, every construct was tested in a small scale for expression in 2xTY-media. Appropriate competent cells were freshly transformed with the expression plasmid and the cell suspension was plated onto TYE agar plates containing necessary antibiotics. Plates were incubated for 12-16 hours at 37° C. 5 ml of 2xTY-medium was then inoculated with 2-4 colonies of the BL21 transformants and incubated at 37° C in a shaking incubator at 210 rpm. Typically, expression was induced at an OD₆₀₀ of 0.4-0.5. A sample of 1 ml uninduced cells was pelleted (13000 rpm, 5 min, 22° C benchtop), resuspended in 100 µl PGLB and boiled immediately for 2 min before inducing protein expression in the remainder of the culture by addition of IPTG to a final concentration of 0.2 - 1 mM. 1 ml samples of the induced cultures were taken after 60, 120 and 180 min and centrifuged (13000 rpm, 5 min, 22° C, benchtop), cells resuspended in 100 µl PGLB, boiled immediately for 2 min and analyzed by SDS polyacrylamide gel electrophoresis.

2.3.2 Large scale expression

Expression of larger protein quantities was achieved by increasing the *E. coli* culture volume to 6, 12 or 24 liters with 500 ml culture in each of the 2000 ml flasks. As for the small

scale expression, appropriate competent cells were freshly transformed with the expression plasmid and the cell suspension was plated onto TYE agar plates containing necessary antibiotics. Plates were incubated for 12-16 hours at 37° C. The next morning, a starter culture (or 2 cultures) of 100 ml 2xTY-medium was inoculated with several transformed colonies and incubated at 37 °C with constant shaking (210 rpm). When the OD₆₀₀ was 0.4, each flask with 2xTY-medium was inoculated with 5 ml of the starter culture. The cultures were incubated at 37° C with constant shaking (210 rpm) until the OD₆₀₀ was 0.35 - 0.45. Two samples of 1 ml uninduced cells from two different flasks were pelleted (13000 rpm, 5 min, 22° C, benchtop), resuspended in 100 µl PGLB and boiled immediately for 2 min. Protein production was then induced by addition of IPTG to a final concentration of 0.2 - 1 mM and the cultures were incubated further for 2.5-4 h. The cells were finally collected by centrifugation (6000 rpm, 6 min, 22° C, RC26, SLA3000). The cell pellets were transferred to 50 ml polypropylene tubes, shock frozen in liquid nitrogen and stored at -20° C until further use.

2.3.3 Solubility test

To test the solubility of the overexpressed protein, a sample taken from a test expression cell culture, pelleted and resuspended in a bacterial lysis buffer. The cells were incubated on ice for 10 min before sonication (2-3 pulses of 5 s) using the micro probe of the sonicator at level 5. 50 µl of the whole cell extract were mixed with 50 µl PGLB and boiled immediately for 2 min. 500 µl of the whole cell extract were pelleted (13000 rpm, 5 min, 22° C, benchtop). The pellet was resuspended in 500 µl PGLB and boiled for 2 min. 50 µl of the supernatant were mixed with 50 µl PGLB and boiled for 2 min. Equal volumes of the whole cell extract, the pellet fraction and the supernatant were analyzed by SDS-PAGE.

2.4 Insect cell (baculo) expression methods

The insect cell (baculo) expression system offers an advantage over bacterial systems when the protein (usually eukaryotic) cannot be overexpressed or properly folded in bacteria. It is a relatively fast way to produce posttranslationally modified proteins in quantities relevant for structural studies.

The Sf21 cell line (derived from the pupal ovarian tissue of *Spodoptera frugiperda*) is highly susceptible to infection with *Autographa californica* nuclear polyhedrosis virus (AcNPV baculovirus). This expression method is based on the production of the recombinant baculovirus, with the gene of interest placed under a strong viral promoter (polyhedrin or p10), which is used to infect Sf21 cells in suspension. The overview of the baculovirus expression system is shown in the Figure 2.1 and procedures are described in detail in the following sections.

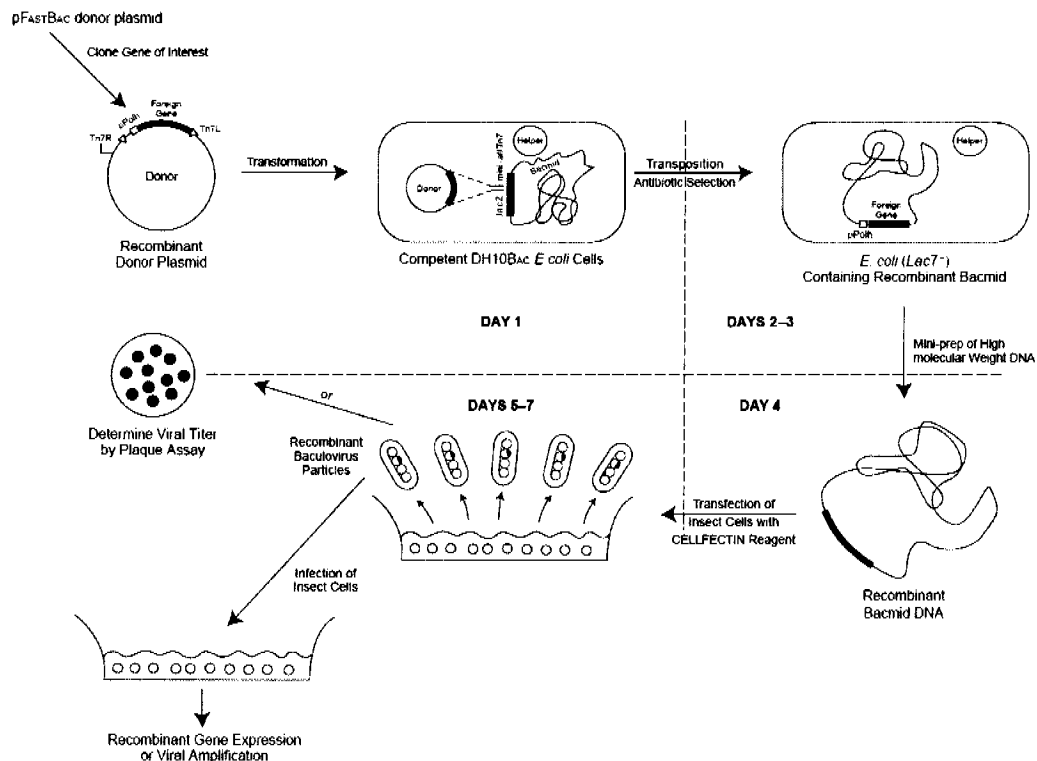


Figure 2.1. The Bac-to-Bac baculovirus expression system (Invitrogen). The scheme shows the steps in the production of the initial virus stock. (The figure is taken from the Bac-to-Bac manual.)

2.4.1 Initiating a cell culture from a frozen stock

First step in insect cell culturing is initiation of the cell culture from a frozen stock of cells. A vial of frozen Sf21 cells was removed from the liquid nitrogen storage dewar, thawed quickly and centrifuged. 10% DMSO-containing medium was removed and the cells were mixed with 40 ml SF-900 II SFM media in a sterile 50 ml polypropylene tube. The cells were then poured into a 150x200 mm sterile tissue culture dish which was incubated for 2 h at 27° C to let the cells attach to it. The medium was removed gently and the new medium was added. The cells were then incubated for 4 days with regular checking under the light microscope, and then removed from the culture dish by repeated pipetting action. Afterwards they were transferred to a 50 ml Erlenmeyer flask and cultured in suspension on a shaking incubator at 90 rpm. When the cell doubling occurred roughly every 20 h, the cells were ready for use.

2.4.2 Maintaining a cell culture in suspension

Sf21 cells were kept at the cell density between 0.5×10^6 and 8×10^6 cells/ml, ideally around 1×10^6 cells/ml. Densities below 0.5×10^6 cells/ml are not recommended because cells will divide more slowly, and densities above 4×10^6 cells/ml were avoided since the cells in this case are too dense to receive good aeration. The cell density was measured using the Neubauer counting chamber and the light microscope. Since their doubling time is approximately 20 h, the cells were diluted using the SF-900 II SFM every 24-48 h. The volume of the cell culture was usually between 1/20 and 1/5 of the shaker flask volume, to avoid cell drying-out (with smaller volumes) and bad aeration (with bigger volumes). Cells from one initial stock were cultured and passaged for approximately 3 months, since then the cell viability started to decrease, as seen from the visual examination using the light microscope. Healthy, viable cells had a round shape, while infected or old cells had even double size or developed different irregular shapes with a lot of cell debris present in the culture.

2.4.3 Production of recombinant bacmid DNA

The gene to be expressed was first cloned into a plasmid transfer vector (based on pFastBac) using the subcloning methods described above. The transfer vector was then used for

the transformation of the competent DH10MultiBac *E. coli* cells using a standard transformation protocol. The DH10MultiBac cells contain a baculovirus shuttle vector (bacmid) which contains a mini-F replicon, a kanamycin resistance marker, and a segment of DNA encoding the *lacZα* peptide from a pUC based cloning vector. The N-terminus of the *lacZα* gene has an insertion containing the attachment site for the bacterial transposon Tn7 (which does not disrupt the reading frame of *lacZα*). This bacmid propagates in DH10MultiBac *E. coli* cells like a large plasmid that can complement a *lacZα* deletion present in the bacterial genome to form blue colonies in the presence of chromogenic substrate (X-gal) and IPTG as an inducer. DH10MultiBac cells also contain the helper plasmid that encodes for Tn7 transposase complex, which catalyzes the transposition reaction of the expression cassette from the transfer vector into the target site on the bacmid. The transformation reaction was incubated at 37° C for 8 h in a shaking incubator, to allow transposition to occur. After the incubation, four serial dilutions of the cells were streaked out on agar plates containing kanamycin, gentamycine, tetracycline, IPTG and X-gal. The plates were incubated at 37° C for 24 to 48 h. The transposition disrupts the expression of the LacZα protein making positive colonies white. Eight white colonies were picked and restreaked on the same type of agar plates, to confirm they are positive.

Four confirmed white colonies were inoculated in 2 ml of the LB medium supplemented with kanamycin, gentamycine and tetracycline. After 24 h incubation, the following adapted plasmid prep was used to isolate the high molecular weight bacmid DNA (~130 kb). 1.5 ml of the culture was centrifuged (13000 rpm, 5 min, 22° C benchtop) and cells were resuspended in 0.3 ml of solution 1 (15 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNaseA). 0.3 ml of solution 2 was added (0.2 N NaOH, 1%SDS) and the sample was gently mixed followed by a 5 min incubation. 0.3 ml of 3 M potassium acetate pH 5.5 was added afterwards; the sample was mixed and incubated for 10 min on ice. The sample was then centrifuged (13000 rpm, 10 min, 22° C benchtop) and supernatant transferred to a tube with 0.8 ml isopropanol. The sample was gently mixed and incubated on ice for 10 min. After centrifugation (13000 rpm, 15 min, 22° C benchtop), the supernatant was removed and 0.5 ml of the 70% ethanol was added to the pellet. The sample was mixed again by inverting the tube and centrifuged (13000 rpm, 5 min, 22° C benchtop). All the following steps, including the transfection of Sf21 cells, were done in the sterile hood (at 27° C) to prevent insect cell culture contamination. The supernatant was removed and the pellet dried for 5 min. Bacmid DNA was dissolved in 20 µl of sterile TE buffer.

2.4.4 Transfection of Sf21 cells

Before transfection with the isolated bacmid, 1×10^6 cells were seeded on a 35 mm well (6-well plate) in 2 ml of SF-900 II SFM medium. The cells were allowed to attach to the plate by 30 min incubation. After the wash with the fresh 2 ml of SF-900 II SFM, the cells were ready for transfection. Solutions of 10 μ l of bacmid DNA in 100 μ l of SF-900 II SFM (solution A), and 6 μ l of Cellfectin in 100 μ l of SF-900 II SFM (solution B) were combined, mixed gently and incubated for 30 min. For each transfection, 0.8 ml of SF-900 II SFM was added to the solution AB and overlaid on the Sf21 cells. Bacmid DNA from each of the 4 initial transfections, coming from 4 different white colonies, was used for two identical 6-well plates. One well was used only for the non-transfected cells (as a negative control). The cells were incubated for 5 h, followed by the removal of the transfection mixture and addition of 2 ml of fresh SF-900 II SFM. The cells were incubated for 3-4 days. Then the virus (present in the media overlaying the cells) was harvested and stored at 4° C in the dark. Since a very low titer virus was expected after the initial transfection ($\sim 2 \times 10^7$ pfu/ml), it was necessary to proceed with the virus amplification step before the virus was used for preparative protein production.

2.4.5 Virus amplification and maintenance

As the virus titer, as well as the total amount of virus obtained from the initial transfection is not sufficient for protein production, the virus had to be amplified. This was done by infecting a 20 ml suspension of Sf21 cells at the density of 1×10^6 cells/ml with 2 ml of the initial virus. The infection was monitored daily by observing cell density and shape. When the cell density in the suspension reached 4×10^6 cells/ml, the cells were diluted back to 1×10^6 cells/ml. This process was continued as long as cells were dividing. Once the cell number was constant, the virus was harvested after a 72 h incubation. This adapted procedure allowed fast amplification of the initial virus giving a high titer virus sufficient for large scale protein expressions.

2.4.6 Plaque assay for virus titer determination

A plaque assay was carried out for accurate determination of the virus titer. A 6-well plate was seeded with 1.5×10^6 cells in 2 ml of SF-900 II SFM and the cells were allowed to attach for

1 h. Several dilutions of the virus (10^{-4} to 10^{-8}) were prepared, each in 1 ml of the medium. The supernatant was removed from the 6-well plate and the wells were inoculated with 0.9 ml of the diluted virus solution. One well was used only for the non-infected cells (as a negative control). The plate was incubated for 1 h. During this time, agarose for overlaying the cells was prepared by dissolving 0.6 g of the low-melting point agarose in 20 ml of the medium. The solution was sterilized and temperature was adjusted to 40-42° C. The virus inoculum was aspirated off the cells and 3 ml of the agarose medium solution (at 28-37° C) was added to each well. After the agarose had hardened, 1 ml of the medium was added and the plate was incubated for 5-7 days at 27° C. The medium was then discarded and 1 ml of the dye solution (1 ml of 0.33% neutral red stain in 12.5 ml of the medium) was added. The plate was incubated for 3 to 24 h. Afterwards, the plaques were counted and virus titer determined.

2.4.7 Expression tests

Before starting preparative protein expression, viruses had to be tested for protein production. This was done by infecting 20 ml cell suspensions (at the 2×10^6 cells/ml density) in several 250 ml Erlenmeyer flasks by serial dilutions of the virus. The protein expression was analyzed by SDS-PAGE 3 days after the infection. The smallest virus volume that stops the cell division immediately after addition and gives highest protein expression was determined. This relative volume was subsequently used for preparative protein production.

2.4.8 Small and medium scale expression in shaker flasks

For the small (50-200 ml) or medium scale (400-2000 ml) protein expression, Sf21 cells were prepared in adequate Erlenmeyer flasks at the density of 2×10^6 cells/ml. After addition of the optimal virus volume (as determined in the expression tests), the cells were incubated for 3 days. After confirming the protein expression by SDS-PAGE, the cells were harvested by centrifugation (2000 rpm, 5 min, 27° C, tabletop). The cell pellets were flash frozen and stored at -20° C. The supernatant, containing the virus, was also kept and stored at 4° C, if the expression test using this virus showed positive results.

2.4.9 Large scale expression in the bioreactor

The large scale protein expression (10 L) in Sf21 cells was done in the bioreactor. The bioreactor was prepared by autoclaving it twice with 5 L of water. It was allowed to cool down until it reached about 30° C. The cooling/heating system was turned on and all of the other connections attached (temperature sensor, oxygen electrode, mixing motor, peristaltic pump, aeration tubing). After the temperature stabilized at 27° C, the oxygen electrode was calibrated and the control loops were set as described in the instrument manual. The water was removed from the bioreactor into the Erlenmeyer flask in the sterile hood using a peristaltic pump. 8 L of SF-900 II SFM was pumped into the bioreactor, followed by 2 L of Sf21 cells, 5×10^6 cells/ml density. The cells were incubated for 1-2 days to achieve the density of 2×10^6 cells/ml. The high titer virus used for the cell infection was filter sterilized and subsequently pumped into the bioreactor. The amount of the virus was determined previously in small scale expression tests, and then scaled up to the bioreactor volume.

The condition of the cells in the bioreactor was monitored every 12 h by taking a 1 ml sample through a septum on a sample loop tubing. The protein expression was analyzed by SDS-PAGE. Approximately 3 days after virus infection, the cells were harvested by using the peristaltic pump. They were centrifuged and stored as described in the previous section.

The bioreactor was disconnected and cleaned several times by pumping water through the main chamber and all the tubing. Finally, it was washed with 10% acetic acid, rinsed with water and autoclaved twice with 5 L of water.

2.4.10 Preparation of nuclear and cytosolic insect cell extracts for SDS PAGE analysis

Sf21 cell pellets were resuspended in the “insect cell lysis buffer” (100 μ l for 1×10^6 cells) and centrifuged (4000 rpm, 5 min, 22° C, benchtop). The supernatant (“cytosolic fraction”) contained the soluble, cytosolic proteins. The pellet, containing the protein aggregates and the nuclei, was washed with lysis buffer and then resuspended in the same volume of the “insect cell nuclei soak buffer”. After 3-4 h incubation by rotation at 4° C, the sample was centrifuged again (4000 rpm, 5 min, 22° C, benchtop). Supernatant (“nuclear soak fraction”) contained the soluble nuclear proteins that could be soaked out of the nuclei. The pellet, containing insoluble proteins

and those that can not be soaked out of the nuclei, was washed with the soak buffer and resuspended again in the same volume. This sample was sonicated for 3-4 s and centrifuged again (12000 rpm, 5 min, 22° C, benchtop). The supernatant (“nuclear fraction”) contained the soluble nuclear proteins and the pellet (“insoluble fraction”) the insoluble proteins. All the fractions were subsequently analyzed by SDS-PAGE.

2.5 Protein purification methods

2.5.1 Preparation of protein extract (cell lysis)

Exact preparation of the crude cell protein extract from which a certain protein was to be purified depended on the cell source (bacterial or Sf21 insect cells) and properties of the protein prepared.

2.5.1.1 Bacterial cell lysis

For bacterial protein extract preparation, the cells from a large scale expression, pelleted and shock frozen, were thawed in a 37° C water bath. The thawed cells were mixed with bacterial lysis buffer (or other, if specified) in a volume approximately 10x the original cell pellet volume and incubated on ice for 10 min. Lysis was obtained by sonication with 5-10 pulses of 10 s using the largest probe of the Sonicator at level 6-10. The obtained whole cell extract was centrifuged (10000 rpm, 20 min, 4° C, RC26, SLA3000) and the pellet and supernatant were separated and stored on ice.

2.5.1.2 Preparation of insect cell cytoplasmic and nuclear fraction

For insect cell protein extract preparation, the cell pellets were thawed at room temperature and resuspended in 10x cell pellet volume of the insect cell lysis buffer (or other, if specified). The cells were homogenized using the manual homogenizing device and centrifuged (4000 rpm, 15 min, 22° C, tabletop) to separate the cytosolic fraction from the nuclei. This procedure was repeated twice to ensure complete cell lysis. The pellet contained the nuclei and supernatant represented the cytosolic fraction. According to the localization of the protein in question, purification was either continued with the cytosolic fraction, or the nuclear fraction was prepared as follows. The nuclei were washed twice with the lysis buffer, resuspended in the nuclear soak buffer (or other, if specified) in 10x of the nuclei volume. The nuclei were mechanically lysed by passing them two times through the cell cracker and centrifuged in the preparative ultracentrifuge or Sorvall RC 26 to separate the insoluble fraction (pellet) from the soluble nuclear extract (supernatant).

2.5.2 Batch protein purification

Batch purification method was sometimes used during the protein purification protocol development to increase the speed of purification, test different purification conditions or save the sample. The chosen chromatography resin was placed into an Eppendorf tube in the volume of 50-200 μ l (Falcon tubes were used for bigger volumes). The resin was pre-equilibrated with the buffer by mixing and centrifuging (2000 rpm, 2 min, 22° C, benchtop/tabletop). The supernatant was carefully removed without disturbing the resin. Crude protein extract was added to the resin and the tube was incubated rotating at 4° C for 30 min. The unbound sample was separated by centrifugation, the resin was washed and the protein eluted using the appropriate buffer. After the elution, the resin was washed with the elution buffer and mixed with the SDS-PAGE loading buffer. After boiling for 5 min, it was cooled down and loaded for SDS-PAGE along with the other fractions to determine if some sample remained bound after elution.

2.5.3 Fast protein liquid chromatography (FPLC)

FPLC was carried out with the conventional chromatography equipment (peristaltic pump), with a Gilson system or with an ÄKTA FPLC system. The protein solution to be purified was centrifuged using the preparative ultracentrifuge or Sorvall RC 26 centrifuge. The supernatant was loaded onto a pre-equilibrated column and eluted with a specific gradient or an appropriate buffer. The peak fractions as detected at 280 nm were analyzed by SDS-PAGE and pooled. The pooled peak fractions were either dialyzed at 4° C in MWCO 12-14 kDa dialysis bags against dialysis buffer or used directly.

2.5.4 Production of TEV protease

TEV protease (containing a His-tag fusion) was expressed from the pET19b-TEV plasmid in BL21 (DE3+pArg) *E. coli* strain obtained from the Nenad Ban's laboratory at the ETH Zurich. 4 L of transformed cells were grown in 25 μ g/ml kanamycin and 100 μ g/ml ampicillin 2xTY media at 30° C until OD₆₀₀ of 0.7. The temperature was then reduced to 20° C and cells were induced with 0.1 mM IPTG for 5 h. The cells were harvested by centrifugation (5000 rpm, 10 min, 4° C, SLA-3000).

The cell pellet was resuspended in 100 ml of the lysis buffer (25 mM sodium phosphate-NaOH pH 7.5, 300 mM NaCl, 1 mM MgCl₂, 10% glycerol, 0.05% Tween-20, 2 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF). The cells were broken using the cell cracker, passing the suspension twice through the apparatus which was previously cooled on ice. The cell debris was separated from the soluble protein extract by a preparative ultracentrifuge run (40 000 rpm, 60 min, 4° C, Ti-70).

The supernatant was filtered through a 0.2 µm syringe filter and loaded onto a 5 ml TALON™ column (low pressure Pharmacia column, 2.5 cm in diameter) pre-equilibrated in lysis buffer. The column was then washed with about 10 column volumes of the lysis buffer and 10 column volumes of the wash buffer (25 mM sodium phosphate-NaOH pH 7.5, 300 mM NaCl, 1 mM MgCl₂, 10% glycerol, 0.05% Tween-20, 2 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF, 25 mM imidazole-HCl pH 7.5). The TEV protease was eluted from the column using a gradient which runs from 0 to 100% of the elution buffer (25 mM sodium phosphate-NaOH pH 7.5, 500 mM NaCl, 1mM MgCl₂, 10% glycerol, 0.05% Tween-20, 2 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF, 500 mM imidazole-HCl pH 7.5) over 60 ml.

Immediately after the elution, which occurs between 335 and 440 mM imidazole, the peak fractions containing the TEV protease were collected and diluted 1:1 with the dialysis buffer (25 mM sodium phosphate-NaOH pH 7.5, 500 mM NaCl, 1 mM MgCl₂, 10% glycerol, 1 mM DTT, 0.5 mM EDTA, 1 mM PMSF, 25 mM imidazole pH 7.5). Dialysis was performed over night at 4 deg, in the volume of 500 ml of the dialysis buffer.

After dialysis, the sample was mixed with an equal volume of 2x storage buffer (50 mM Tris-HCl pH 7.0, 1 M NaCl, 20% glycerol, 2 mM DTT, 1 mM EDTA) concentrated in the Amicon concentrating device to 2 mg/ml. The equal volume of the 100% glycerol was then added and the sample was stored at -80° C. The activity was retained for over 2 years. (Note: On the SDS-PAGE the TEV protease appears as a double band due to an internal cleavage.)

The final yield was about 12 mg of purified protein per 1 L of the cell culture.

2.5.5 Production of core human TBP

Core human TBP (TBPC) was expressed from the pET28aTBPC plasmid constructed during this work from pET3aTBPC by standard subcloning (using *Xba* I and *Bam*H I), to change

the antibiotic resistance gene from ampicillin to kanamycin and ensure plasmid stability. This plasmid was used for transformation of *E. coli* Rosetta strain since codon bias prevents high expression levels in *E. coli* strains without additional tRNA genes (especially tRNA^{Arg}). A 150 ml preculture was grown overnight in 2xTY media supplemented with 25 µg/ml kanamycin. 120 ml of this preculture was used for inoculation of 12 L of SPM media supplemented with 25 µg/ml kanamycin. The cells were grown at 37° C until OD₆₀₀ reached 0.5 (this step took about 10 h, since the cells grow very slow in SPM media). The cells were induced with 0.2 mM IPTG and harvested after 3 h by centrifugation (5000 rpm, 10 min, 4° C, SLA-3000).

The cell pellet from 6 L of cell culture (previously frozen and thawed in the 37° C water bath) was resuspended in 300 ml of the TG300 buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 1 mM benzamidine, 10 mM beta-mercaptoethanol and 10% glycerol). The cells were broken using the sonicator, with 3x 10 s pulses in 50 ml Falcon tubes placed on ice. The cell debris was separated from the soluble protein extract by a preparative ultracentrifuge run (40 000 rpm, 60 min, 4° C, Ti-70).

The first purification step was done at 4° C. Supernatant was loaded onto a 10 ml heparin sepharose column (low pressure Pharmacia column, 2.5 cm in diameter) pre-equilibrated in TG300 buffer. The column was washed with TG300 buffer until the UV absorbance reached a low plateau. The wash was continued with the TG300 buffer supplemented with 2 M urea until the lower absorbance plateau was reached again, which normally required about 10 column volumes. A linear gradient from 300 mM NaCl to 1500 mM NaCl formed with a Bio-Rad 385 gradient former (40 ml TG300 and 40 ml TG1500) was applied after another TG300 wash (10 column volumes). TBPC was eluted in the middle of the gradient, together with a 10 kDa contaminant and several others at very low concentrations (as analyzed on the SDS-PAGE). The middle of the main peak was collected into a 150 ml Erlenmeyer flask, transferred to dialysis tubing and dialyzed against 1 L of H200 (2 mM HEPES-NaOH pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 10 mM 2-mercaptoethanol) for 15 h at 4° C, with one change of the buffer after 3 h.

After dialysis, the sample was centrifuged (20000 rpm, 10 min, 22° C, SS-34) to remove the white precipitate. The supernatant, containing all of TBPC, was loaded onto a CM-3SW (2.15 cm diameter x 15 cm length) column, mounted on the Gilson purification system and pre-equilibrated in H200, using a 50 ml superloop for the injection. The following run was set up at

room temperature: 20 ml H200, gradient of 30-60% H1000 (2 mM HEPES-NaOH pH 7.5, 1 M NaCl, 0.5 mM EDTA, 10 mM 2-mercaptoethanol) over 160 ml, 35 ml H1000.

The sample eluted at 550 mM NaCl in a linear gradient between 440 mM and 680 mM NaCl, separate from the main 10 kDa contaminant eluting at 200 mM NaCl (as analyzed on the SDS-PAGE). The peak fractions (containing less than 2% of 12 kDa and 15 kDa contaminants) were pooled and dialyzed against 2 L of BG100 (2 mM BisTris-HCl pH 7.0, 100 mM KCl, 0.1 mM EDTA, 0.5 mM DTT and 10% glycerol) at 4° C for 15 h with one buffer exchange after 3 h. The dialyzed sample was centrifuged (20000 rpm, 10 min, 4° C, SS-34) and the supernatant concentrated to about 1 ml using the Amicon concentrating device. Glycerol was added to a total concentration of 20%. The sample was centrifuged (13000 rpm, 10 min, 4° C, benchtop) and dispensed into the Eppendorf tubes in 200 µl aliquots. The aliquots were flash frozen in liquid nitrogen and stored at -80° C.

The final yield was about 2 mg of purified protein per 1 L of the cell culture.

2.5.6 Production of human TAF8/10 complex

The TAF8/10 purification protocol (developed by I. Berger) was optimized during this work. TAF8/10 complex was co-expressed in Sf21 insect cells from a pFBDMTAF8TAF10 vector-derived virus (containing a C-terminal His-tag fusion on TAF8). 1 L of Sf21 cells at density of 1×10^6 cells/ml were infected with a high titer virus and harvested after 3 days by centrifugation (2000 rpm, 5 min, 27° C, tabletop). The expression of the complex was analyzed by SDS-PAGE. The cell pellet was resuspended in 40 ml of the lysis buffer (50 mM HEPES-NaOH pH 7.6, 100 mM KCl, 0.1 % NP-40, 10 mM 2-mercaptoethanol, 1 mM PMSF, 10 mM imidazole), homogenized and centrifuged (4000 rpm, 15 min, 4° C, tabletop). Supernatant (cytosolic fraction), containing the TAF8/10 complex, was removed. The pellet was resuspended and centrifuged again to extract the rest of the complex. Supernatants were finally combined and centrifuged in the preparative ultracentrifuge (40000 rpm, 60 min, 4° C, Ti-70).

The supernatant was loaded onto a 5 ml TALON™ column (self-made low pressure Pharmacia column, 2.5 cm in diameter), pre-equilibrated in the lysis buffer. All the following purification steps were done using the Aekta purification system, at 4 deg. The column was washed with buffer A (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM 2-mercaptoethanol, 1

mM PMSF, 10 mM imidazole) until the absorbance reached a low plateau (about 20 column volumes), and then with wash buffer (25 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM 2-mercaptoethanol, 1 mM PMSF, 10 mM imidazole) for 10 column volumes. After another wash with the buffer A (5 column volumes), the complex was eluted using the linear gradient between the buffer A and the elution buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM PMSF, 200 mM imidazole), over 50 ml.

The peak fractions were pooled and concentrated to about 5 mg/ml using the Amicon concentrating device. Concentrated sample was centrifuged (12000 rpm, 5 min, 4° C, benchtop) and loaded onto a S200 gel filtration column, pre-equilibrated in the S200 buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 1 mM PMSF). Only 0.5 ml of the sample was used for one run. The complex eluted after 11.5 ml, possibly indicating a tetrameric oligomerization state of the complex. SDS-PAGE analysis of the fractions was done to confirm the separation of the complex from the high molecular weight contaminants. Peak fractions were pooled and concentrated to 5 mg/ml. Glycerol was added to the sample to the final concentration of 30%. The sample was aliquoted, flash-frozen in liquid nitrogen and stored at -80° C.

The final yield of the TAF8/10 purification was about 10 mg protein per 1 L of the cell culture.

2.6 Protein analysis methods

2.6.1 Determination of protein concentrations

Protein concentrations were determined by UV-absorption spectroscopy. The absorbance (A) from 320 nm to 220 nm was measured against a corresponding buffer control sample. Protein concentrations (c_{Protein}) were calculated with the formula $(A_{280} - A_{320}) = c_{\text{Protein}} * \epsilon * d$ using extinction coefficients (ϵ) calculated with the program ProtParam (on ExPASy server, <http://www.expasy.ch/tools/protparam.html>). The extinction coefficients used in this work are listed in the Table 2.1.

Protein/complex	Extinction coefficient at 280 nm ($\text{cm}^{-1} \text{mg}^{-1} \text{ml}$)
TBpc	0.435
full-length TAF1	0.851
TAF1-construct 2	0.649
full-length TAF2	1.129
full-length TAF2 TEV mutants (single)	1.133
full-length TAF2 TEV mutants (double)	1.136
TAF2 core domain	1.288
TAF2 tail domain	0.318
TAF8/10	0.415
TEV	1.146

Table 2.1. Extinction coefficients of the proteins used in this work.

2.6.2 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analyzed using denaturing 12% (for TAF1 and TAF2) or 15% (for other proteins) separating gel (60:1 acrylamide:bisacrylamide, 0.75 M Tris-HCl pH 8.8, 0.1% SDS), covered with a 5% stacking gel (60:1 acrylamide:bisacrylamide, 120 mM BisTris-HCl pH 6.8,

0.1% SDS). The gels were made using the BioRad gel electrophoresis system. Samples were diluted with an equal volume of PGLB and boiled for 2 min before loading. Gels were run for 45-60 min at 10 W using 50 mM Tris (pH not adjusted), 0.1 M glycine, 0.1 % SDS as running buffer. After SDS-PAGE, protein bands were visualized by Coomassie staining (Laemmli, 1970).

2.6.3 Western blotting

SDS-PAGE of protein samples was performed as described. The gel was removed from the electrophoresis apparatus and soaked for 5 minutes in 30 ml of the Western blot buffer. A PVDF membrane (7.5 cm x 7.5 cm, 0.2 μ m, BioRad) was pre-soaked in methanol for 15 min, and then in Western blot buffer for 5 min, along with the sponges and two pieces of 3 MM paper. The blotting sandwich was assembled in the following order: gray bottom support (to face negative electrode), a sponge, a piece of 3MM paper, equilibrated gel, PVDF membrane, a piece of 3MM paper, a sponge, clear top support (to face positive electrode). The assembly was transferred to the blotting apparatus, which was filled with the Western blot buffer, and the proteins were transferred to the membrane by electrophoresis at 200 mA for 90 minutes at 22° C or overnight at 4° C. The gel was Coomassie-stained to determine the amount of remaining unblotted protein. The PVDF membrane was stained with Ponceau Red for 5 min, washed with water and equilibrated in 30 ml of 1x TBS buffer for 5 minutes. It was transferred then to 50 ml of the TBSM buffer (1x TBS supplemented with 20 mg/ml of milk powder) for 30 minutes, and incubated in 20 ml of the TBSM buffer containing the appropriate primary antibody (1:1000 dilution) for 1 h at 22° C. After three wash steps (5 min each) with 50 ml 1x TBSM, it was incubated in 20 ml TBSM buffer containing the anti-mouse/goat-IgG-alkaline phosphatase secondary antibody for 1 hour at 22° C (1:1000 dilution). Finally, the filter was washed 3 times for 5 minutes in fresh 50 ml 1x TBS and transferred to 10 ml 1x TBS pH 9.5 containing 100 μ l of the NBT/BCIP substrate for visualization. After the color was developed, the reaction was stopped with deionized water. The membrane was dried over night between two 3 MM papers.

2.6.4 N-terminal sequencing

The proteins or protein fragments to be sequenced were separated by SDS-PAGE. A PVDF membrane and the blotting apparatus were prepared as described for the Western blotting analysis. Electroblothing was performed either at 50 V for 2 hours at 22° C or overnight at 20 V, 4° C. The PVDF membrane was stained with fresh and filtered Coomassie staining solution at room temperature for 2-3 min, and then destained with absolute ethanol (HPLC grade) for 5 min. The membrane was air-dried at room temperature. The appropriate bands were cut out with a sterile scalpel and submitted for N-terminal sequencing by Edman degradation to the Protein Service Laboratory of the Functional Genomic Center Zürich.

2.6.5 Mass spectrometry

Samples for the mass spectrometry (MALDI-TOF) analysis were submitted to the Protein Service Laboratory of the Functional Genomic Center Zürich. They were prepared in an appropriate buffer, containing less than 5% glycerol and not more than 200 mM salt. The concentration was between 0.1 and 1 mg/ml, in the case of hTAF2 proteolytic fragments.

2.6.6 Limited proteolysis

As a first step in a limited proteolysis experiment, it was necessary to determine the appropriate protease concentration to be used. The protein to be analyzed was purified and prepared at the concentration of 1-2 mg/ml in a low salt buffer. It was then incubated for 30 min at room temperature with different dilutions of the protease (trypsin, chymotrypsin, elastase or subtilisin), usually at 1:100 to 1:10000 protease:protein mass ratios. The reaction was stopped by adding PGLB to the sample and immediate flash-freezing in liquid nitrogen. The resulting protein fragments were analyzed by SDS-PAGE. The protease concentration that produced fragments in the required molecular mass range was chosen for the experiment.

The time-course limited proteolysis experiment was performed by using the protein at the 1-2 mg/ml concentration with the previously determined concentration of the protease. The reaction was set up in an Eppendorf tube, in a volume of 150 µl, at room temperature. After protease was added, the 20 µl samples were taken out into a 0.5 ml Eppendorf tube, mixed with

PGLB and flash-frozen, typically after 1, 5, 15, 30, 60 and 120 min. The protein fragments were subsequently analyzed by SDS-PAGE.

For preparative work, the reactions were multiplied (and not scaled up), for example by using 10 Eppendorf tubes with a 150 μ l reaction volume. The reactions were stopped by adding 3 mM PMSF, followed by centrifugation (12000 rpm, 5 min, 22° C, benchtop).

2.6.7 Electrophoretic mobility shift assay

Electromobility shift assays (EMSA) or band shift experiments were performed to analyze the DNA binding activity of TAF2 (Fried and Crothers, 1981; Garner and Revzin, 1981). The protein was incubated on ice for 15-30 min with the oligonucleotide in a specific bandshift buffer. Concentrations of protein and DNA were kept either in the nanomolar or micromolar range. The sample was loaded onto a prerun polyacrylamide gel (usually 10% (40:1 acrylamide:bisacrylamide) in 0.5x TBE buffer, 18 cm x 18 cm x 1 mm). The gel was run at 4° C for 4-6 h at a constant voltage of 150 V with constantly recirculating buffer. The bands corresponding to the protein-DNA complexes and the free DNA were visualized either by autoradiography for bandshifts in the nanomolar range, or using a UV transilluminator after staining with ethidium bromide solution for 5 min for bandshifts in the micromolar range.

2.6.8 Analytical ultracentrifugation (sedimentation velocity experiments)

Sedimentation velocity experiments with the full-length TAF2 protein were carried out using a Beckman XL-I analytical ultracentrifuge equipped with scanner optics. The initial sample absorption at 280 nm was around 0.5. Samples (400 μ l) were equilibrated in 0.1-HEMG buffer (25 mM HEPES-NaOH pH 7.6, 0.1 M KCl, 0.5 mM EDTA, 5 mM MgCl₂ and 10% glycerol) in the analytical ultracentrifuge chamber under vacuum for 30-60 min at 20° C prior to sedimentation at 40 krpm in 12 mm double sector cells and an 8-hole rotor. Scans were collected at 280 nm in a continuous scan mode using a 0.001 cm radial step size. Boundaries were analyzed by the continuous $c(s)$ distribution method using the program Sedfit. Sedimentation coefficients were corrected to $s_{20,w}$ using a partial specific volume of 0.7378 ml/g for the full-length TAF2 protein ($M_w=137$ kDa) and density and viscosity specific to the 0.1-HEMG buffer

(1.02762 and 2.0071, respectively). Processed data were plotted as boundary fraction versus $s_{20,w}$ to yield the integral distribution of sedimentation coefficients. For calculation of mean sedimentation coefficients, only the dominant homogeneous fraction of the boundary was taken into account. All sample points include 2 independent measurements.

2.7 Circular dichroism (CD) spectroscopy

The phenomenon of circular dichroism (CD) comes from the fact that proteins, as chiral molecules, absorb left and right circularly polarized light with different extinction coefficients. The CD spectrometer measures the difference in the intensities of left and right circularly polarized light after passing through the sample and expresses this as the ellipticity Θ , the angle measuring the extent to which the transmitted light is elliptical. Since the difference between the two components is usually very small, on the order of 0.0001, the ellipticity is only a few 1/100th of a degree.

The molar mean residue ellipticity $[\Theta]_{\text{MRW}}$ of a protein can be obtained by dividing the measured molar ellipticity by the number of amino acids in the protein (n):

$$[\Theta]_{\text{MRW}} = [\Theta] / n \quad (\text{Unit: Deg cm}^2 \text{ dmol}^{-1})$$

The CD spectra of proteins with different secondary structure components are characteristically distinct in the far-UV range of 180-250 nm (Figure 2.2). Proteins that only contain α -helices as regular secondary structure show a strongly negative CD signal with minima at 208 and 222 nm, while all- β -sheet proteins have a weaker negative signal with a minimum at 215 nm. Since the far-UV CD spectra of native and unfolded proteins are generally very different (the minimum of unfolded proteins is below 200 nm), far-UV CD spectroscopy is an excellent tool to measure folding and conformational transitions in proteins.

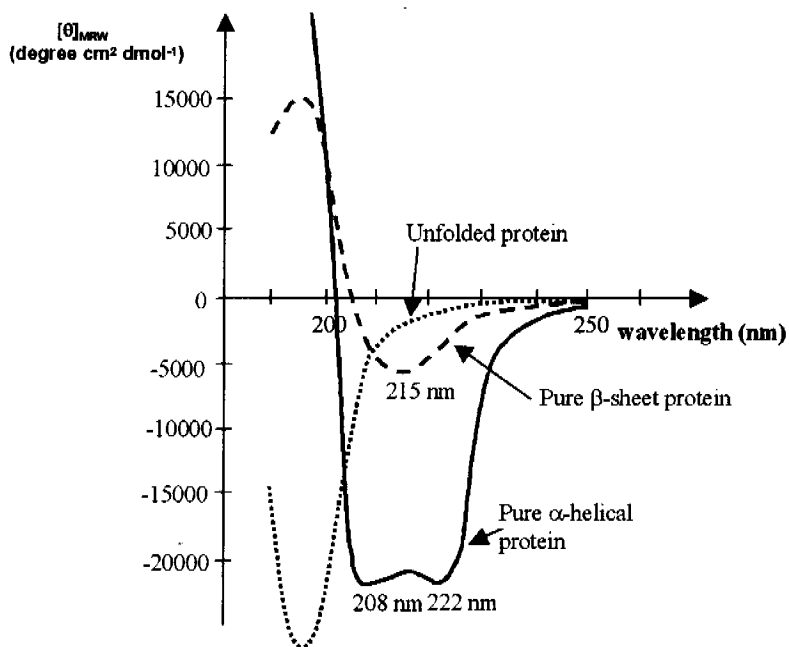


Figure 2.2. Typical far-UV CD spectra of unfolded, all- α and all- β proteins.

2.7.1 Thermal stability measurements

When a protein containing alpha helices denatures due to the increasing temperature, it loses its secondary structure and the CD signal at 220 nm is largely increased. Melting temperature (T_m) of proteins is estimated from the dependence of the measured CD signal at this wavelength to the temperature using the following equation:

$$\text{Observed Signal} = \frac{((S_N + m_N \cdot T) + (S_U + m_U \cdot T) \cdot \exp(\Delta H_m / (R \cdot T_m) \cdot (T - T_m) / T_m))}{(1 + \exp(\Delta H_m / (R \cdot T_m) \cdot (T - T_m) / T_m))}$$

$$m_U = 0.00001; S_U = -8; m_N = 1e-5; S_N = -10; \Delta H_m = -300000; R = 8.3145;$$

where

T is the temperature in K,

R is the gas constant,

m_U is dependence of signal of unfolded protein on temperature,

S_U is signal of unfolded protein (intercept point at 0 K),

m_N is dependence of signal of native protein on temperature,

S_N is signal of the native protein (intercept point at 0K),

T_m is the melting temperature in K and

ΔH_m is enthalpy change for unfolding at T_m .

The measured data is fitted to this equation using the SigmaPlot software. If met with 'singular coefficient matrix' error (error due to the 'non-proper' starting values), the value of S_N can be varied from -10 to -60.

The instrument settings used for the T_m measurements were the following: 220 nm wavelength, 25 - 80° C temperature range, 1° C/min temperature slope, 100 mdeg sensitivity and 1 nm bandwidth. Measurements were performed with 500 μ l of the sample at the concentration of 0.2 mg/ml in a clean 0.1 cm cuvette. T_m was calculated from two independent measurements.

2.7.2 Measurements of secondary structure content

For far-UV CD spectra that are used to measure the secondary structure content of proteins, it is critical to use buffers that absorb only weakly in this wavelength range. Table 2.1 gives a series of solutions and their absorbance in the far-UV CD range. If high ionic strength is required for protein stability and solubility, it is best to use NaF or KF instead of NaCl and KCl to maintain ionic strength, because the chloride anion absorbs significantly in the far-UV range.

CD spectra measurements performed in the course of this work were done in CD buffer (10 mM Tris-Cl pH 8.0, 100 mM KF, 2 mM $MgCl_2$ and 2 mM 2-mercaptoethanol), into which the sample was extensively dialyzed. 200 μ l of the sample at 0.1 to 0.3 mg/ml concentration was placed in a clean 0.1 cm cuvette and the CD signal was measured in the JASCO CD spectrometer. The following parameters were used: 250-190 nm wavelength, 100 mdeg sensitivity, 0.2 nm data pitch, 50 nm/min scanning speed, 0.2 s response and 10 accumulations per spectrum. All spectra were measured at 20° C. The data was analyzed using the KaleidaGraph™ software. The CD signal of each spectrum was corrected by subtracting the buffer signal, and then converted into the molar mean residue ellipticity $[\Theta]_{MRW}$ using the following formula:

$$[\Theta]_{MRW} (\text{Deg cm}^2 \text{ dmol}^{-1}) = (C * 100 * Mw) / (\gamma * 0.1 * n * 1000)$$

where C is the measured CD signal, M_w the molecular weight of the protein (in Da), γ the mass concentration (in mg/ml) and n the number of amino acids.

Compound	no absorbance above	A of a 10 mM solution in a 1 mm cuvette at			
		210 nm	200 nm	190 nm	180 nm
NaF, KF	170 nm	-	-	-	-
NaCl	210 nm	-	0.02	>0.5	>0.5
MES-NaOH pH 6.0	230 nm	0.07	0.29	0.29	>0.5
HEPES pH 7.0	230 nm	0.24	0.89	2.12	?
Cacodylic acid-NaOH pH 6.0	210 nm	0.01	0.2	0.22	?
NaH ₂ PO ₄	195 nm	0	0	0.01	0.15
Na ₂ HPO ₄	210 nm	-	0.05	0.3	>0.5
MOPS-NaOH pH 7.0	230 nm	0.1	0.34	0.28	>0.5
Tris-H ₂ SO ₄ or Tris-HCl pH 8.0	220 nm	0.02	0.13	0.24	>0.5
Boric acid-NaOH pH 9.1	200 nm	-	-	0.09	0.3
GdmCl pH 8.0	230 nm	0.009	0.064	0.07	?
DTT (reduced form)	250 nm	1.25	3.75	4.25	4.25
NaH ₂ PO ₄ /Na ₂ HPO ₄ pH 6.5	220 nm	0.003	0.05	0.125	?
EDTA pH 8.0	260 nm	1.5	2	4	?
Glycerol (1% = 135.7 mM)	300 nm	0.007	0.035	0.82	?

Table 2.2. Buffers and salts in far-UV CD measurements.

2.8 Surface plasmon resonance (*BiaCore*) methods

Surface plasmon resonance (SPR) is a quantitative method for investigation of molecular interactions (Karlsson, 2004). One molecule ("ligand") is immobilized by chemical coupling to a sensor chip surface, and the interactions are monitored by injecting samples of the other molecule ("analyte") over this surface. The association and dissociation are monitored and recorded in real time. The sensor surface may be regenerated between injections by selective dissociation of the interaction partners. Regeneration solutions ensure complete dissociation, without affecting the binding characteristics of the immobilized partner.

SPR can be used either for obtaining qualitative data, to prove two molecules are interacting, or quantitative data, to establish the dissociation constants and the kinetics of the interaction.

2.8.1 The principle of the surface plasmon resonance (SPR)

The principle of the surface plasmon resonance is shown in Figure 2.3. When light is reflected at an interface between two media of different refractive index, light coming from the side with the higher refractive index is partly reflected and partly refracted. Above a certain critical angle of incidence, no light is reflected across the interface and total internal reflection is observed. Although the incident light is totally reflected, an electromagnetic field component called the evanescent wave penetrates a short distance into the medium of lower refractive index. If the interface between the media is coated with a thin layer of gold, and light is monochromatic and polarized, the intensity of the reflected light is significantly reduced at the specific angle of incidence, producing a sharp "shadow". This phenomenon is called the surface plasmon resonance (SPR), and the angle at which the shadow is observed is the SPR angle. The SPR signal depends on the refractive index of the medium into which the evanescent wave propagates, on the low refractive index side of the surface. This refractive index is affected by the surface concentration of the solutes. Therefore, monitoring of the SPR angle provides a real time measure of changes in the surface concentration. For example, if one molecule is bound to the surface and another passed over it, the SPR signal will be shifted. This change is recorded in real time as SPR signal and expressed in resonance units (RU). One RU corresponds to 0.0001° , and for most proteins this is a change in concentration of $1\text{pg}/\text{mm}^2$ on the sensor surface. The

signal is measured continuously to form a sensogram (Figure 2.4), which provides a complete record of association and dissociation of the interactants (BIAtechnology handbook, 1994).

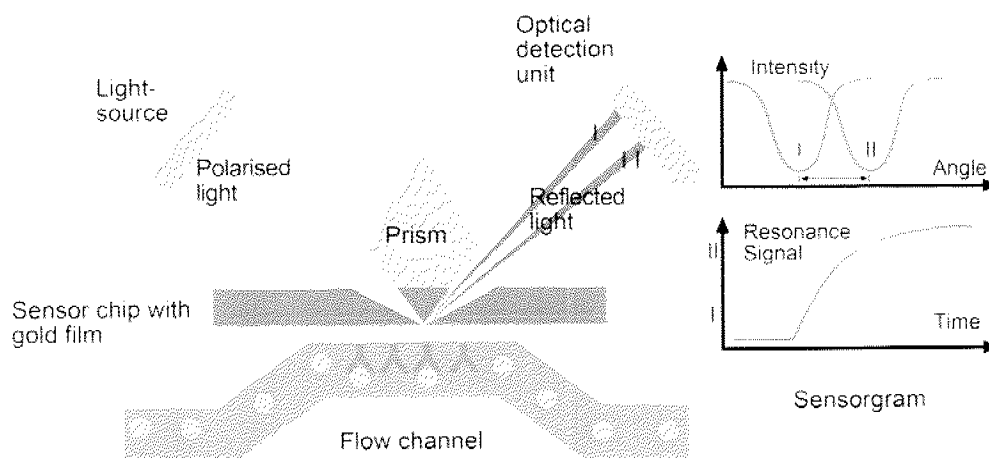


Figure 2.3. The principle of the surface plasmon resonance (SPR).

SPR arises when light is reflected under certain conditions from a conducting film (gold layer) at the interface between 2 media (sample and the glass of the sensor chip) of different refractive index. SPR causes a reduction in the intensity of reflected light at a specific angle of reflection. This angle varies with the refractive index close to the surface on the side opposite from the reflected light on the sample side. (The figure is taken from the BiaCore presentation slides.)

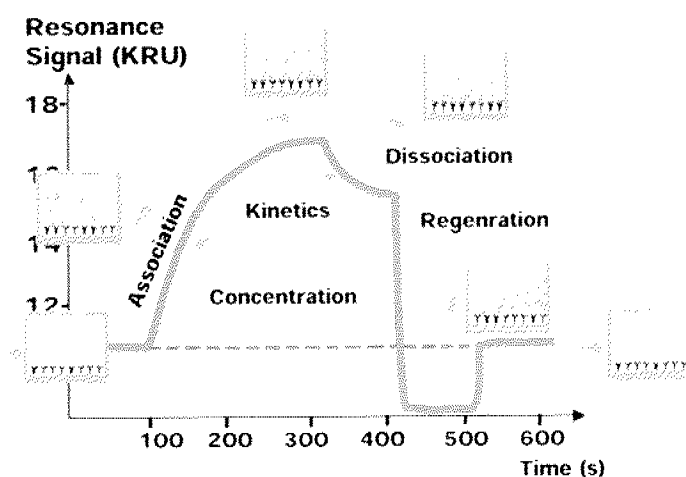


Figure 2.4. A scheme of a sensogram.

The sensogram provides quantitative information in real-time on specificity of binding, active concentration of molecule in a sample, kinetics and affinity. (The figure is taken from the BiaCore presentation slides.)

2.8.2 BiaCore sensor chips

For an interaction study, one of the interaction partners is immobilized onto the sensor surface of a BiaCore sensor chip. Immobilization occurs by direct coupling to the surface or via a suitable molecule already coupled to the surface. Sensor surface is chosen according to the nature of the molecule to be coupled and the requirements for the analysis.

The most commonly used sensor chip is CM5, with a matrix of carboxymethylated dextran covalently attached to a gold surface. Molecules are covalently coupled to the sensor surface via amine, thiol, aldehyde or carboxyl groups. CM5 has a high binding capacity which potentially gives a high response. CM4 is a variant of this chip, with only 30% of CM5 binding capacity due to the smaller number of carboxyl groups.

C1 is a matrix-free carboxymethylated sensor chip with a low binding capacity. The flat surface allows interactions to take place closer to the surface. This might be advantageous when working with very large molecules or molecular complexes and for those interactions that may be affected by the presence of dextran.

SA chip has a matrix of carboxymethylated dextran pre-immobilized with streptavidin. It is suitable for immobilization of biotinylated interaction partners (like nucleic acids) and enables orientated immobilization by controlled ligand biotinylation.

Before the immobilization takes place, the ligand has to be pre-concentrated on the chip surface. The ligand is prepared in a buffer with a pH below its pI value, so the net charge of the protein is positive. Since the chip surface is negative due to the carboxyl groups, most of the protein molecules can come close enough to the surface for the immobilization reaction to occur.

2.8.3 Experimental setup

Before performing a BiaCore experiment, one has to choose the appropriate ligand, sensor chip, immobilization method and the amount of the immobilized ligand. Additionally, one has to make sure the chip surface can be regenerated and the unwanted binding reduced to the minimum.

The molecule chosen for the ligand should be as pure as possible, stable in the regeneration conditions, specific for the analyte (in case it is in the complex mixture) and, if possible, smaller than the analyte in size (the response is proportional to the size). The sensor

chip and the immobilization method are chosen according to the properties of the ligand and the measured interaction. The most common method is amine coupling of proteins to CM5 chip surface, since all proteins have free amine groups for coupling to the surface.

The amount of the immobilized ligand is chosen according to the type of measurements performed. For kinetic measurements it should be as low as possible to avoid the formation of vertical analyte gradient, called the “mass transport effect”. The amount of ligand to be immobilized was estimated using this equation:

$$R_{\max} = (\text{Mw analyte} / \text{Mw ligand}) \times R1 \times S, \text{ where}$$

R_{\max} is the maximum binding capacity of the analyte,

R1 is the amount of ligand immobilized on the chip, and

S is the stoichiometry of the binding reaction.

Generally, for the kinetic analysis, the ligand was immobilized to the levels which give R_{\max} of 100-500 RU, and for the steady state binding analysis, higher immobilization levels were used.

The running buffer and regeneration buffer had to be determined experimentally for each different ligand/analyte combination. The running buffer was designed such that it maximally reduces the non-specific binding of the analyte to the chip surface, as monitored in the reference lane containing no immobilized ligand. The regeneration buffer had to be designed in a way that it achieves complete dissociation of the analyte, while preserving the activity of the ligand.

All experiments were performed at 25° C using two flow cells on the chip. One flow cell was the reference cell, and the other contained the immobilized ligand. The reference cell is used as a control to monitor non-specific binding to the chip surface. The analyte was prepared in the running buffer to minimize the bulk effect (change in response due to refractive index difference between the running buffer and analyte solution) and centrifuged before injection (12000 rpm, 5 min, RT, benchtop).

2.8.4 Protein immobilization by amine coupling

For coupling of the protein to the chip surface via amine groups, the CM5 or CM4 chip is first activated with a 1:1 mixture of NHS:EDC (N-hydroxysuccinimide:N-ethyl-N'-(dimethylaminopropyl)-carbodiimide), to give reactive succinimide esters. These esters can react spontaneously with amines and other nucleophilic groups on the ligand to form covalent links. NHS:EDC mixture was prepared fresh and immediately used for an 8 min injection on the chip surface at a flow rate of 5 μ l/min. As a second step, the ligand is injected over the chip surface at the same flow rate, until the desired immobilization level is achieved. The final step is inactivation of the succinimide esters by an ethanolamine solution, injected for 10 min at the flow rate of 10 μ l/min. All solutions were prepared from the BiaCore amine coupling kit according to the manual.

2.8.5 Protein immobilization by thiol coupling

For coupling of the protein to the chip surface via thiol groups, one first has to make sure that the protein contains free cysteine side chains for the reaction. The CM5 or CM4 chip is activated with a 1:1 mixture of NHS:EDC, as described for the amine coupling. A fresh solution of the 80 mM PDEA (2-(2-pyridinyldithio)ethanamine) is prepared in 0.1 M sodium borate buffer (pH 8.5). This solution is injected on the chip surface for 4 min at the flow rate of 5 μ l/min, in order to introduce reactive disulfide groups. The ligand is then injected in a buffer containing no reducing agents (DTT or 2-mercaptoethanol) at the same flow rate, until the desired immobilization level is achieved. The final step is inactivation of the excess reactive groups by 50 mM cysteine and 1 M NaCl in 0.1 M sodium acetate buffer (pH 4.0). This solution is injected for 4 min at the flow rate of 10 μ l/min. All solutions were prepared from the BiaCore thiol coupling kit according to the manual. The running and regeneration buffers used with the thiol-coupled ligand had to be free of reducing agents.

2.8.6 Biotinylated DNA immobilization

The biotinylated DNA oligonucleotides were immobilized on streptavidin-coated chips (SA chips). The chip surface was first activated by three 1 min injections of 50 mM NaOH,

followed by injections of 0.5 M NaCl. The biotinylated DNA molecules were injected stepwise in small volumes until the desired immobilization level was achieved.

2.8.7 Data analysis

The sensograms produced during the SPR experiments were analyzed using the BIAevaluation 3.2 software package (BiaCore). Only the curves obtained by subtraction of the inactive lane (background signal) from an active lane were used for the analysis. The subtraction curves were adjusted by a Y-transformation to align the baselines (the resonance units produced by the immobilized molecule) and subtraction of the zero-concentration analyte. The steady state analysis was performed as described in the BIAevaluation handbook, using the following equation:

$$R_{eq} = K_a * c * R_{max} / (K_a * c * (n+1)); \quad K_d = c \text{ at } 50\% R_{max}$$

where

R_{eq} is the response of the ligand to the analyte at equilibrium;

R_{max} is the maximum response of the ligand to the analyte at equilibrium;

K_a is the equilibrium association constant of ligand and analyte;

K_d is the equilibrium dissociation constant of ligand and analyte;

c is the molar concentration of the analyte;

n is the number of analyte molecules interacting with the ligand.

The data was fitted according to the 1:1 Langmuir binding model using the BIAevaluation software (version 3.2). χ^2 values and residual plots were used to assess the quality of the fit.

2.9 Crystallization screens and methods

2.9.1 General remarks

Material to be crystallized (human TAF2 and its shorter constructs) was prepared fresh, concentrated to 5-10 mg/ml and centrifuged (13000 rpm, 15 min, 4° C, benchtop) immediately before setting up drops. All solutions used for crystallization were prepared from chemicals of the highest available quality (e.g. Fluka microselect if available) using MilliQ water, and filtered through 0.2 µm filters (Sartorius). Crystallization plates were incubated either at 22° C or at 4° C, in special boxes that absorb vibrations. The progress of the crystallization was monitored weekly under a microscope equipped with a polarization filter. For the Cartesian setups, the plates were stored in the Crystallization plate hotel (Discovery Partners) and monitored automatically on a fixed time schedule.

2.9.2 Vapor diffusion crystallization

Crystallization setups were prepared according to the vapor-diffusion method in the sitting drops setups, using the 12 x 8 multi-well plates sealed with 'Crystal clear' tape. Equal volumes (1-2 µl) of sample and precipitant were mixed in the middle chamber of a 12 x 8 multi-well plate. The droplets were equilibrated against 100 µl of reservoir solution containing the appropriate dilution of precipitant in single wells sealed with a tape.

2.9.3 Crystallization using the Cartesian robot

The Cartesian robot (operated by Beat Blattman, University of Zürich) was used for the majority of the crystal setups. This robot is able to produce 3 crystallization drops per well and uses Greiner 96 well plates. The plates are prepared with commercial or custom screens using the Lissy liquid handling robot which pipettes 100 µl of the crystallization buffer into each well. The plate is then transferred into the Cartesian robot, which mixes 50-200 nl of the protein solution with 50-200 nl of the crystallization buffer.

2.9.4 Crystallization screens

A wide range of the crystallization screens, available at the Cartesian robot crystallization facility, was used for the initial crystallization setups:

Grid Screen	Grid Screen Description
GS001	PEG4000/0.2M Li ₂ SO ₄ //PEG8000 5% - 30%, pH gradient 3.0 ~ 10.0
GS002	(NH ₄) ₂ SO ₄ @ 4°C in % max. solubility, pH gradient 3.0 ~ 10.0
GS003	PEG4000/0.2M Li ₂ SO ₄ //PEG6000 5% - 30%, pH gradient 3.0 ~ 10.0
GS005	PEG4000//PEG8000 5% - 30%, pH gradient 3.0 ~ 10.0
GS006	PEG4000//PEG6000 5% - 30%, pH gradient 3.0 ~ 10.0
GS007	Clear Strategy Screen™ I & II; pH 5.5 & 6.5
GS008	Clear Strategy Screen™ I & II, pH 7.5 & 8.5
GS009	Custom Grid-Screen; 8 Buffer Solutions vs. 12 Concentration-Steps of a Precipitant (PEG10K: 10-25%, MES-HCl 0.1 M pH 6.0 & 6.5 – designed during this work)
GS013	PEG 10000 5% - 30%; PEG 5000 MME 5% - 30%; pH gradient 3.0 ~ 10.0
GS015	Sigma-Fluka Factorial Screen #82009, Crystallization Basic Kit and #70437, Crystallization Extension Kit for Proteins. (Identical to Hampton Screen)
GS018	Sigma-Fluka #75403, Crystallization Cryo for Proteins and Sigma-Fluka #86684, Crystallization Low Ionic Kit for Proteins
GS020	PEG400 15-45%//PEG4000 5-30%, pH gradient 4.5 ~ 9.4, NaCl 1.0 M + 150 mM (Membrane Protein - MacKinnon Screen)
GS024	PEG400 15-45%//PEG4000 5-30%, pH gradient 4.5 ~ 9.4, ammonium formate; zinc acetate, potassium iodide (Membrane Protein - MacKinnon Screen)
GS031	Structure Screen I & II, HT96; Cat # MD1-30; Molecular Dimensions Ltd.
GS033	Prescreen for estimating the optimal protein concentration for crystallization

In addition, the following custom screen was designed during this work for manual crystallization of human TAF2 (reservoir solution concentration indicated):

	1	2	3	4	5	6
A	50 mM MES-NaOH pH 6.0 9% PEG 3350	50 mM MES-NaOH pH 6.0 13% PEG 3350	50 mM MES-NaOH pH 6.0 17% PEG 3350	50 mM MES-NaOH pH 6.0 21% PEG 3350	50 mM MES-NaOH pH 6.0 25% PEG 3350	50 mM MES-NaOH pH 6.0 27% PEG 3350
B	50 mM MES-NaOH pH 6.0 9% PEG 3350	50 mM MES-NaOH pH 6.0 13% PEG 3350	50 mM MES-NaOH pH 6.0 17% PEG 3350	50 mM MES-NaOH pH 6.0 21% PEG 3350	50 mM MES-NaOH pH 6.0 25% PEG 3350	50 mM MES-NaOH pH 6.0 27% PEG 3350
C	50 mM Tris-HCl pH 8.0 9% PEG 3350	50 mM Tris-HCl pH 8.0 13% PEG 3350	50 mM Tris-HCl pH 8.0 17% PEG 3350	50 mM Tris-HCl pH 8.0 21% PEG 3350	50 mM Tris-HCl pH 8.0 25% PEG 3350	50 mM Tris-HCl pH 8.0 27% PEG 3350
D	50 mM Tris-HCl pH 8.0 9% PEG 3350	50 mM Tris-HCl pH 8.0 13% PEG 3350	50 mM Tris-HCl pH 8.0 17% PEG 3350	50 mM Tris-HCl pH 8.0 21% PEG 3350	50 mM Tris-HCl pH 8.0 25% PEG 3350	50 mM Tris-HCl pH 8.0 27% PEG 3350

2.10 *Bioinformatics tools, databases and software*

Primary structure analyses, such as calculations of DNA and protein parameters, and manipulation of sequences and sequence alignments were performed using the programs DNAMAN, version 5.2.9 (Lynnon BioSoft, Quebec, Canada) and ProtParam (on ExPASy server, <http://www.expasy.ch/tools/protparam.html>). Database searches were performed in Genbank, Swiss-Prot, EMBL and TrEMBL using Web interfaces (e.g. ExPASy or ENTREZ on NCBI). Similarity searches were performed using BLAST (Basic Local Alignment Search Tool) interfaces at NCBI. Sequence alignments were produced using the ClustalW server (<http://ch.embnet.org>), and the formatting of the alignments was done using the ESPript 2.2 (<http://esprict.ibcp.fr>). Secondary structure predictions were made using the HNN prediction method (<http://pbil.ibcp.fr/htm/index.php>) and PSIPRED Protein Structure Prediction Server (<http://bioinf.cs.ucl.ac.uk/psipred>). Predictions of disordered regions were done using the DisEMBL (<http://dis.embl.de>). Protein subcellular localization was predicted by PSORT program (<http://psort.nibb.ac.jp>).

3

RECOMBINANT PRODUCTION AND PURIFICATION OF FULL-LENGTH HUMAN TAF2

3.1 Introduction

In order to obtain milligram quantities of full-length human TAF2 (hTAF2) protein for biochemical characterization and structural studies, it was necessary to produce it recombinantly either in a bacterial or an eukaryotic expression system. Since hTAF2 is a large eukaryotic protein (Mw=137 kDa), it was a good candidate for the expression in Sf21 insect cells (baculovirus system), which provides conditions similar to the human cell. The most important component of this environment is the protein folding machinery and the cellular mechanisms for producing certain post-translational modifications.

It was shown previously (Kaufmann et al, 1998) that native hTAF2 can be purified from the HeLa cell cultures using a naturally occurring stretch of seven histidines near its C-terminus, acting as an internal His-tag (Figure 3.1). This fact was used as a basis for the purification protocol development, which is described in this chapter.

HSDHHHHHHH EHKKKKKKKHK HKHKHKHKHD SKEKDKEPFT FSSPASGRSI RSPSLSD (aa 1140–1197)

Figure 3.1. C-terminal sequence of hTAF2.

The figure shows the last 57 amino acids of hTAF2. Histidines are underlined.

3.2 Expression of hTAF2 in Sf21 insect cells

3.2.1 Virus production and expression tests

The gene for the full-length human TAF2 was obtained from Dr. I. Berger in the pDiFB baculovirus expression vector, under the polh promoter (Figure 3.2). At the N-terminus it contained a FLAG-tag fusion (for detection and/or affinity purification) with an enterokinase cutting site (for possible tag removal).

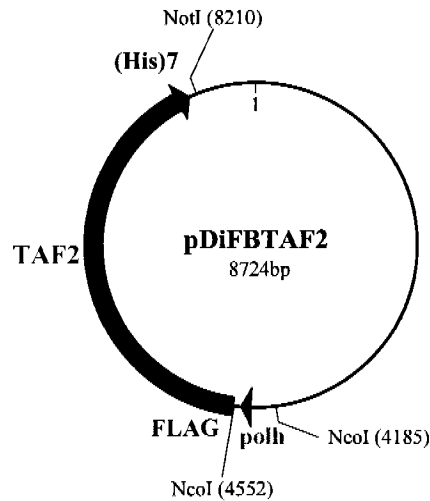


Figure 3.2. Map of the full-length hTAF2 expression vector pDiFBTAF2.

The full-length TAF2 gene in this vector contains a FLAG tag fusion at its N-terminus followed by an enterokinase site. The internal stretch of seven histidines is indicated near the C-terminus.

The recombinant baculovirus containing the hTAF2 gene was produced according to the procedures described in the section 2.4. After the initial transfection, the virus was harvested and the cells were used for SDS-PAGE analysis. The gel didn't show the expression band corresponding to the molecular weight of hTAF2, nor did it show any expression of late viral protein p10 (p10 produces a strong band at molecular weight of 10 kDa in infected cell extracts). The virus amplification was therefore done using 2 ml of the initial virus per well containing 1×10^6 cells. Only after the second amplification the extra band in the expected molecular weight range appeared. The resulting virus (4 ml) was used for infecting 300 ml of the cell culture at the cell density of 1×10^6 cells/ml. The cells stopped dividing after 24 h, and the expression time course was followed during the next 3 days. The additional band, indicating hTAF2 expression, first appeared after 24 h, and reached maximum level after about 72 h (Figure 3.3). The identity of the protein was confirmed by Western blot analysis using the His-tag and Flag-tag antibodies (Figure 3.4).

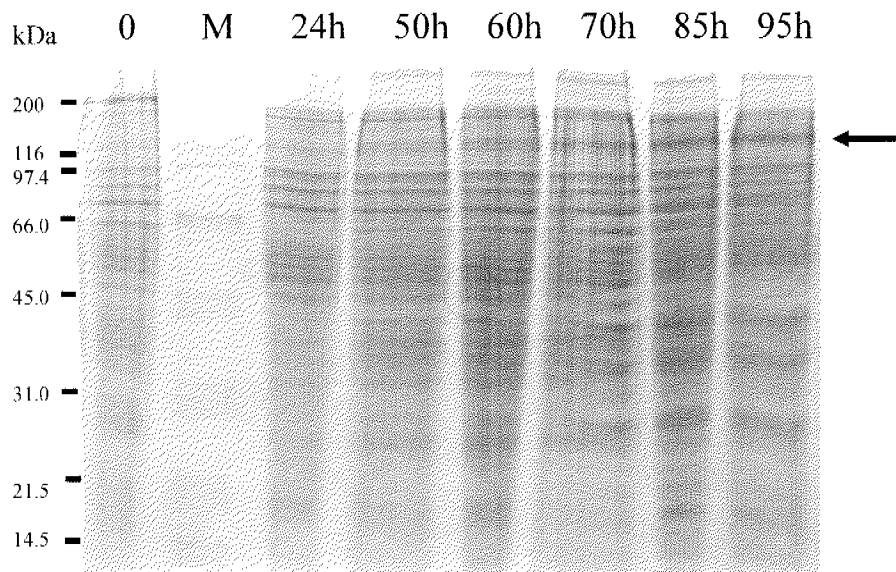


Figure 3.3. Expression time-course of hTAF2 in Sf21 insect cells. The arrow indicates the position of hTAF2 band, appearing after 50 h. Lane “0” contains the negative control sample (cells expressing hTAF1) and lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

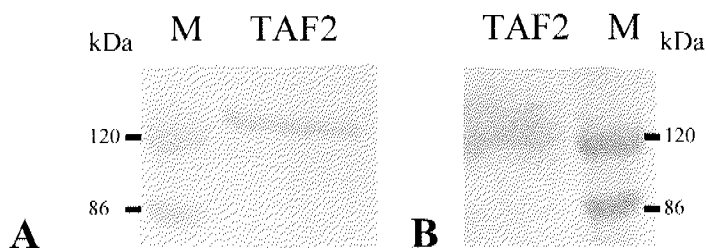


Figure 3.4. Confirmation of hTAF2 identity by western blot. A) Western blot with anti-His antibody. B) Western blot with FLAG-tag antibody. Lane “M” contains the prestained protein marker (BioRad).

All later expressions in shaker flasks (small and medium scale) were done by infection at a density of 2×10^6 cells/ml with the high titer virus capable of infecting all cells immediately. Using higher cell densities caused weak expression, due to the lack of oxygen necessary for intensive protein production.

3.2.2 Large scale expression

Large scale (10 L) expression of hTAF2 was done in the bioreactor, using the procedure described in the section 2.4.9. The bioreactor expression had to be prepared by producing a large volume of the high titer virus, able to infect 10 L of cells at 2×10^6 cells/ml density. The virus was first tested for expression in 20 ml cell cultures grown in shaker flasks. Different virus dilutions were tested, and the one capable of immediately infecting cells at 2×10^6 cells/ml density was chosen. The expression was confirmed by SDS-PAGE. 2 L of 5×10^6 cells/ml culture was prepared and introduced into the bioreactor, already filled with 8 L of media. The cells were left to divide once, producing 10 L of culture at 2×10^6 cells/ml density, and then infected by the virus. The infection was carefully monitored every 12 h by measuring the cell density, visual examination of the cells and oxygen usage. The oxygen levels are measured automatically by the oxygen electrode and plotted by the bioreactor software. 24 h after the cells were infected, their demand for oxygen increased dramatically, indicating the start of the protein production phase. Since the bioreactor is automatically controlling the oxygen levels in the culture, it was possible to obtain good expression levels even with the cell density at 4×10^6 cells/ml. Nevertheless, such high densities were avoided since it involved the risk of not having a virus strong enough to infect high number of cells. In this case, the density could possibly go over the level which allows cell viability (beyond log phase).

48 h after the infection, the virus was harvested from 1 L of the bioreactor cell culture. This virus was tested and stored for future usage. The rest of the cell culture was left in the bioreactor for another 24 h to achieve the maximum protein expression levels. Before harvesting, the expression was confirmed by analyzing cell extracts on the SDS-PAGE. The cells were then harvested, flash-frozen and stored at -80°C .

The expression levels obtained using the bioreactor were comparable to the levels obtained in the small and medium scale expression using the shaker flasks (Figure 3.5). The bioreactor expression, if enough care is taken to avoid contamination of the culture, was less time consuming and therefore more practical than the expression of the equivalent size in shaker flasks. This procedure was done before each large scale hTAF2 purification for crystallization setups.

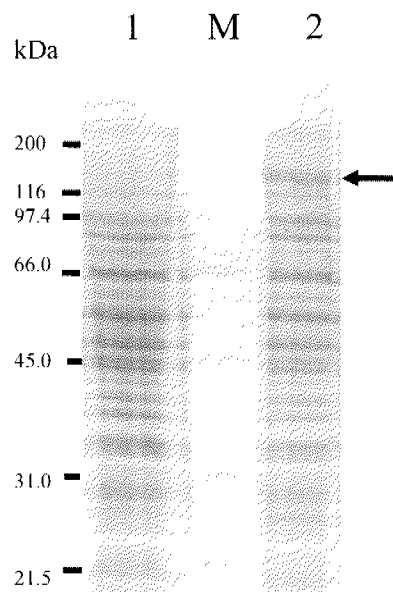


Figure 3.5. Large scale hTAF2 expression in the bioreactor.

Lane 1- negative control (cells expressing hTAF1), lane 2- protein extract from the hTAF2 expressing cells 72 h after infection in the bioreactor. The arrow indicates the position of hTAF2 band. Lane "M" contains the broad range SDS-PAGE protein marker (BioRad).

3.3 Purification of hTAF2

3.3.1 Cytoplasmic localization of hTAF2

To predict the subcellular localization of hTAF2, the protein sequence was submitted to PSORT server (<http://psort.nibb.ac.jp>). Since no nuclear localization signal was found, the protein was predicted to be cytosolic. This result was confirmed experimentally by preparation of the nuclear and cytosolic fraction from Sf21 cells expressing hTAF2 (as described in the section 2.5.1.2), showing that hTAF2 is present mainly in the cytosolic fraction (Figure 3.6).

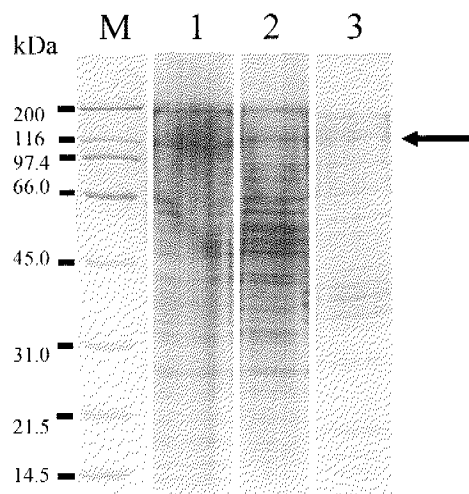


Figure 3.6. Cytoplasmic and nuclear fractions of Sf21 cells expressing hTAF2. Lane 1- whole cell extract, lane 2- cytoplasmic fraction, lane 3- nuclear fraction. The arrow indicates the position of hTAF2 band, present mainly in the cytosolic fraction. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

Preparation of the cytoplasmic fraction was therefore used as the first purification step, allowing the separation of hTAF2 from the nuclear proteins and DNA.

3.3.2 Initial purification trials

The purification of hTAF2, using the cytosolic fraction, was tried initially on a variety of chromatographic resins. The pI of the protein, calculated using the ProtParam tool on the Expasy server (<http://www.expasy.ch/tools/protparam.html>), was predicted to be 8.26, indicating a net positive charge of the protein at pH 7.4 (pH of the insect cell lysis buffer in

which the cytosolic fraction is prepared). Table 3.1 shows the primary structure analysis results and some theoretical parameters of hTAF2 protein sequence.

Number of amino acids:	1197		
Molecular weight (Da):	136928.1		
Theoretical pI	8.26		
Amino acid composition:	Ala (A)	70	5.8%
	Arg (R)	51	4.3%
	Asn (N)	9	4.9%
	Asp (D)	60	5.0%
	Cys (C)	26	2.2%
	Gln (Q)	43	3.6%
	Glu (E)	72	6.0%
	Gly (G)	44	3.7%
	His (H)	51	4.3%
	Ile (I)	58	4.8%
	Leu (L)	116	9.7%
	Lys (K)	88	7.4%
	Met (M)	38	3.2%
	Phe (F)	59	4.9%
	Pro (P)	63	5.3%
	Ser (S)	106	8.9%
	Thr (T)	58	4.8%
	Trp (W)	17	1.4%
	Tyr (Y)	41	3.4%
	Val (V)	77	6.4%

Table 3.1. hTAF2 primary sequence analysis.

As expected, the initial purification trials with the MonoQ (anion exchange) column showed that hTAF2 is not binding at pH 7.5, since it was found in the column flow-through fraction (Figure 3.7). In contrast, hTAF2 could bind to the SP-sepharose (cation exchange) column at pH 7.5 and it could be eluted using a linear salt gradient from 0.1 to 1.0 M KCl (Figure 3.8). The SP-sepharose elution fractions showed a significant enrichment with hTAF2 compared to the cytosolic fraction, but the protein could not be purified to homogeneity using this procedure, or combination of Mono-Q and SP-sepharose columns. hTAF2 also showed binding to the heparin sepharose column, but there was no significant improvement of purity after adding this step to the purification procedure.

Since hTAF2 contains a stretch of seven histidine residues near its C-terminus, it was also a candidate for TALON™ metal affinity purification (Clontech), normally used with His-tag fusion proteins. Initial TALON™ binding trials were done in batch, using 400 µl of the TALON™ resin equilibrated in 0.4-HMG buffer (25 mM HEPES-NaOH pH 7.6, 0.4 M KCl, 10 mM MgCl₂, 10% glycerol, 10 mM 2-mercaptoethanol and 1 mM PMSF). The resin was mixed with 2 ml of the SP-sepharose elution fractions, incubated at room temperature for 5 min, washed with the 0.4-HMG buffer and eluted with 0.4-HMG supplemented with 200 mM imidazole. The SDS-PAGE analysis showed that hTAF2 can bind to the TALON™ resin, improving the purification dramatically (Figure 3.9).

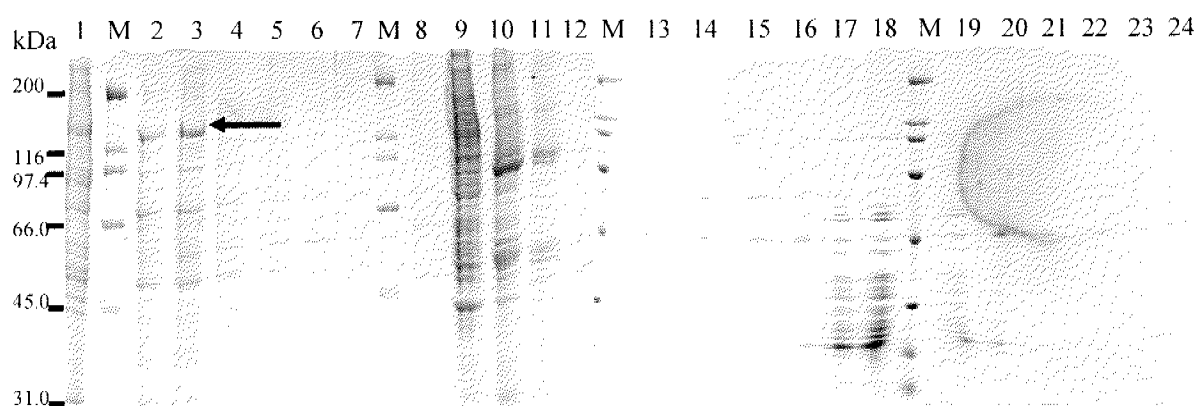


Figure 3.7. Mono-Q column purification of hTAF2 from the cytosolic fraction.

Lane 1- cytosolic extract (column input), lanes 2 to 7- flow-through fractions, lanes 8 to 24- salt gradient elution fractions. Every second fraction was taken for the SDS-PAGE analysis covering all the fractions. The arrow indicates the position of hTAF2 band, present in the flow-through fractions. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

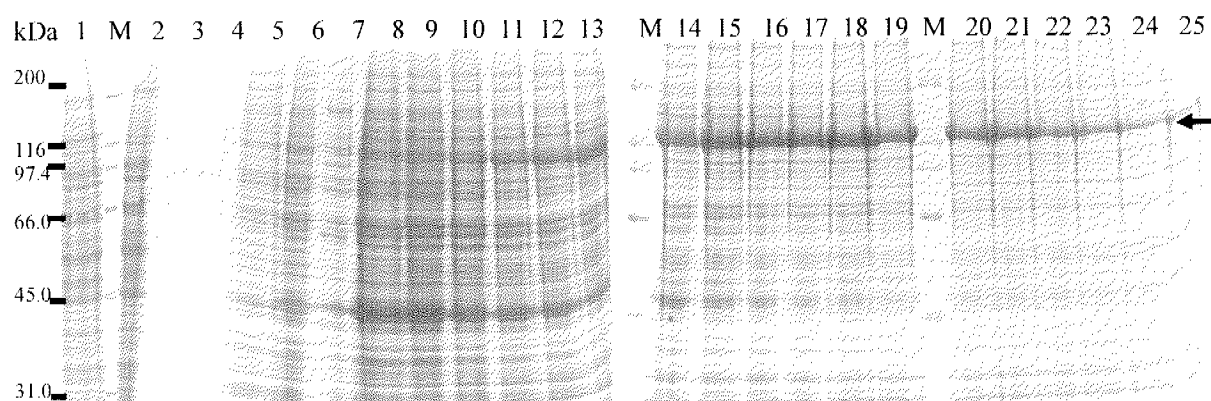


Figure 3.8. SP-sepharose column purification of hTAF2 from the cytosolic fraction. Lane 1- cytosolic extract (column input), lane 2- column flow-through, lanes 3 to 25- 0.1-1.0 M KCl gradient elution fractions. Every or every second fraction was taken for the SDS-PAGE analysis covering all the fractions. The arrow indicates the position of hTAF2 band, present in the flow-through fractions. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

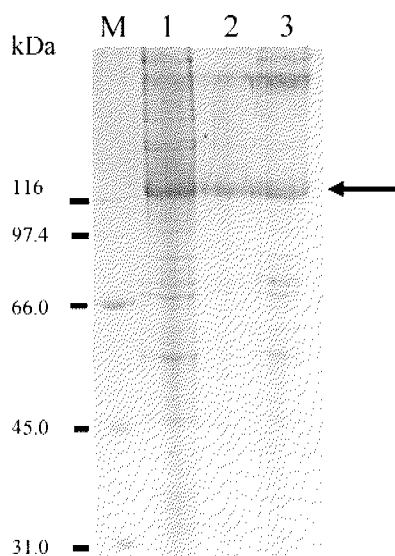


Figure 3.9. TALON™ resin batch purification of hTAF2 with the SP-sepharose column elution fractions. Lane 1- TALON™ resin unbound fraction, lane 2- imidazole elution fraction, lane 3- resin input. The arrow indicates the position of hTAF2 band. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

After the initial purification trials, the purification protocol with the cytosolic fraction was optimized for quality and speed by reducing the number of purification steps, choosing the best purification resin/column for each step according to resolution and capacity, optimizing the buffers and removing the need for buffer exchange by dialysis. The final protocol included four

chromatographic steps, which are described in detail in the following sections and summarized in the section 3.3.8.

3.3.3 Q-sepharose anion-exchange chromatography

For the large scale purification of hTAF2, 30 ml of the cell pellet (previously stored at -20° C) was thawed at room temperature and resuspended in 150 ml of TAF2 lysis buffer (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 50 mM imidazole, 0.1% NP-40, 10 mM 2-mercaptoethanol and 1 mM PMSF). All the subsequent steps were done at 4° C or on ice. The cells were homogenized and centrifuged (4500 rpm, 15 min, 4° C, tabletop). The supernatant (cytosolic fraction) was removed and the procedure was repeated with the pellet. Finally, supernatants were combined and centrifuged to remove the rest of the cell debris and particulate matter (40000 rpm, 60 min, 4° C, Ti-70).

The supernatant, containing hTAF2, was mixed with 30 ml of Q-sepharose resin pre-equilibrated in the TAF2 lysis buffer. The sample was incubated with the resin for 30 min, rotating at 4° C. The sample was centrifuged (2000 rpm, 5 min, 4° C, tabletop) and the supernatant filtered through the miracloth, to remove the resin from the supernatant completely. Since it doesn't bind to the Q-sepharose resin, as established in the initial purification trials with the Mono-Q column (Figure 3.7), hTAF2 was present in the unbound fraction (supernatant).

The Q-sepharose purification step allowed the removal of a large part of the impurities. It was also necessary in order to prevent some proteolytic degradation of the protein during subsequent steps, suggesting that some of the removed impurities were native proteases. This step is performed in batch to be fast and practical; the resin was afterwards extensively washed with the high salt buffer in order to regenerate it for future usage. Afterwards, it was stored in 20% ethanol at 4° C. It could be reused for at least 5 times, without losing the purification effect.

3.3.4 TALON™ affinity chromatography

TALON™ affinity chromatography was done with the Q-sepharose flow-through (unbound) fraction. A self-made 5 ml low-pressure TALON™ column (Pharmacia, 2.5 cm diameter) was pre-equilibrated in the TAF2 lysis buffer. This buffer was adjusted for the optimal

TALON™ purification. It contained 50 mM imidazole (to insure only hTAF2 binds to the resin and elutes more than 95% pure) and no DTT and EDTA, since they are incompatible with metal chelating resins. In the case no imidazole is present during sample loading, a substantial amount of impurities bind to the TALON™ resin, decreasing its capacity for TAF2 and decreasing eluted TAF2 purity (Figure 3.10). Washing the column after sample loading with 50 mM imidazole-containing buffer could remove most of the impurities, but it was not so effective as the presence of 50 mM imidazole in the sample during loading.

The sample was loaded at 4° C using the peristaltic pump, at the speed of 2 ml/min. The column was then washed with 100 ml of 25% TALON™ buffer B (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 200 mM imidazole-HCl pH 7.6, 10 mM 2-mercaptoethanol and 1 mM PMSF) and 75% TALON™ buffer A (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 10 mM 2-mercaptoethanol and 1 mM PMSF). The second wash was performed with 50 ml of the TALON™ wash buffer (25 mM HEPES-NaOH pH 7.6, 1 M KCl, 10 mM MgCl₂, 10 mM 2-mercaptoethanol and 1 mM PMSF), to remove some of the non-specifically bound contaminants. After another wash with 50 ml of 25% TALON™ buffer B and 75% TALON™ buffer A, the following run was started: linear gradient from 25 to 100% TALON™ buffer B over 40 ml, 40 ml 100% TALON™ buffer B. TAF2 eluted in a single peak, starting between 100 and 150 mM imidazole. The column was washed well with TALON™ buffer B and reused 5 times. After fifth time, the decrease in the column capacity was noticed.

The elution fractions, containing TAF2, were analyzed by SDS-PAGE (Figure 3.11) and pooled for the next purification step. The main contaminant was a 55 kDa protein which is present in all TALON™ purifications from insect cells. This unknown protein probably also contains a native histidine stretch with an affinity for the TALON™ resin.

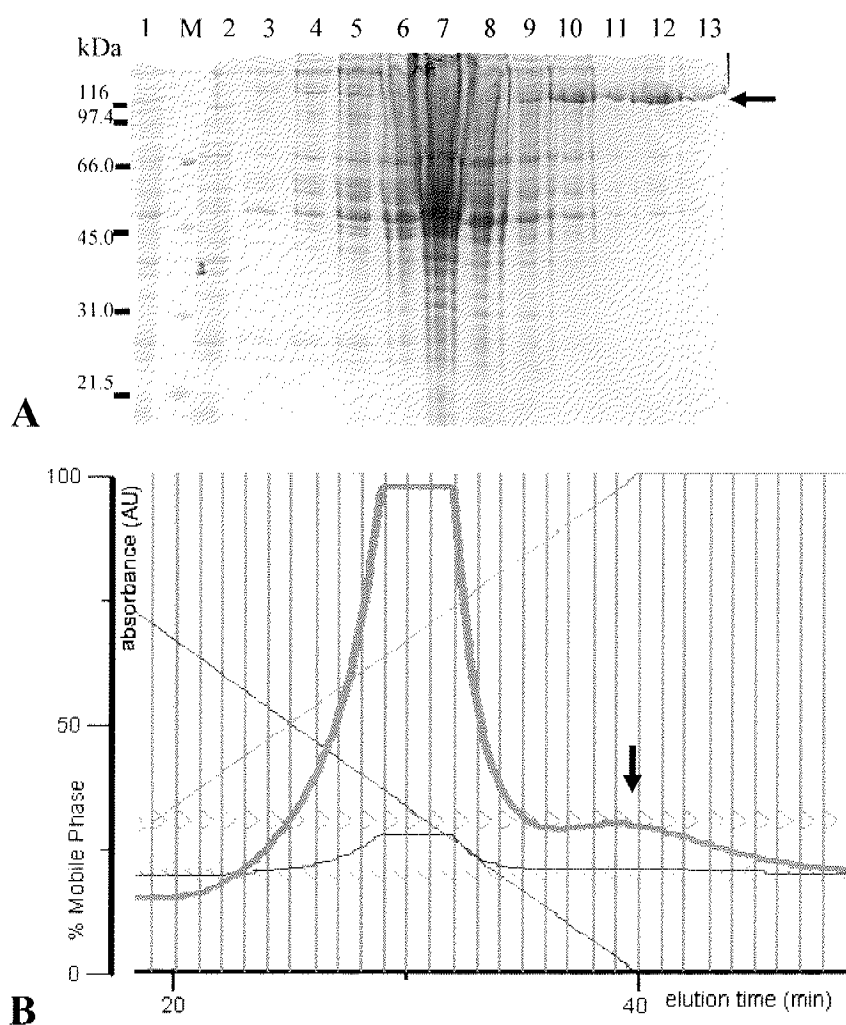


Figure 3.10. hTAF2 TALON™ chromatography with no imidazole-HCl in the loading buffer. A) SDS-PAGE analysis. Lane 1- Q-sepharose flow-through (column input), lane 2- flow-through fraction, lanes 3 to 13- elution fractions, starting from 20 min after the start of the run (see B). Every second fraction was taken for the analysis. The arrow indicates the position of hTAF2 band. Lane “M” contains the prestained protein marker (BioRad). B) Chromatogram of the run. The arrow points to the small peak containing hTAF2.

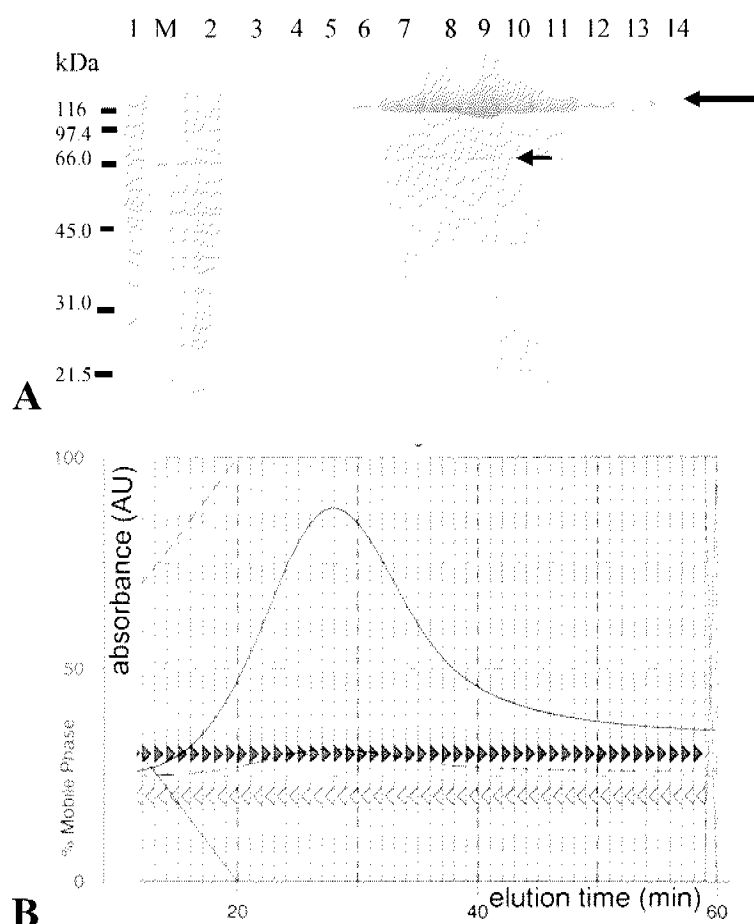


Figure 3.11. hTAF2 TALON™ chromatography with 50 mM imidazole-HCl pH 7.6 in the loading buffer.

A) SDS-PAGE analysis. Lane 1- Q-sepharose flow-through (column input), lane 2- flow-through fraction, lanes 3 to 14- peak elution fractions. The bigger arrow indicates the position of hTAF2 band. The smaller arrow indicates the position of the major contaminant. Lane “M” contains the prestained protein marker (BioRad). B) Chromatogram of the run. The peak contains hTAF2.

3.3.5 Resource S cation-exchange chromatography

For cation-exchange chromatography, a 6 ml Resource S column was chosen because of its high resolution and capacity necessary for large scale purifications. The column was prepared by a 5 column volume wash with Buffer B (25 mM HEPES-NaOH pH 7.6, 1 M KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF), followed by a 5 column volumes wash with Buffer A (25 mM HEPES-NaOH pH 7.6, 0.1 M KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF). All the following steps were done at 4° C. Pooled TALON™ fractions were loaded immediately on the Resource S column using a 50 ml superloop at the speed of 2 ml/min. The column was

protected from any particulate matter by the filter pre-column (Pharmacia). After the sample was loaded, the column was washed with 20 column volumes of Buffer A. The following run was started: linear gradient from 0 to 35% Buffer B over 4 ml, linear gradient from 35 to 65% Buffer B over 40 ml, linear gradient from 65 to 100% Buffer B over 4 ml and 100% Buffer B for 40 ml. TAF2 eluted in the 35-65% Buffer B gradient as a single peak, at approximately 400 mM KCl. The peak fractions were analyzed on the SDS-PAGE (Figure 3.12). Some of the impurities, as visible on Figure 3.12, could be separated since they elute at a somewhat lower ionic strength than TAF2. The main 55 kDa impurity could not be separated completely, as it was visible on a more concentrated sample.

The fractions of Resource S cation exchange purification with no visible contaminants were pooled for the final purification step. When a large scale purification was done, the pooled Resource S fractions were already concentrated enough to load directly on a size-exclusion column. Therefore, this purification step was also a convenient method to change the buffer and quickly concentrate the sample. If no special buffer was required and the sample was not intended for crystallization, it was not necessary to continue with the last purification step. Except saving time, omitting the size-exclusion chromatography allowed a higher yield. Typically 30% of the sample would be lost in the final step due to several reasons: adherence of the protein to the tube, loss through centrifugation, non-specific adherence to the size-exclusion column, and finally, conservative peak collection.

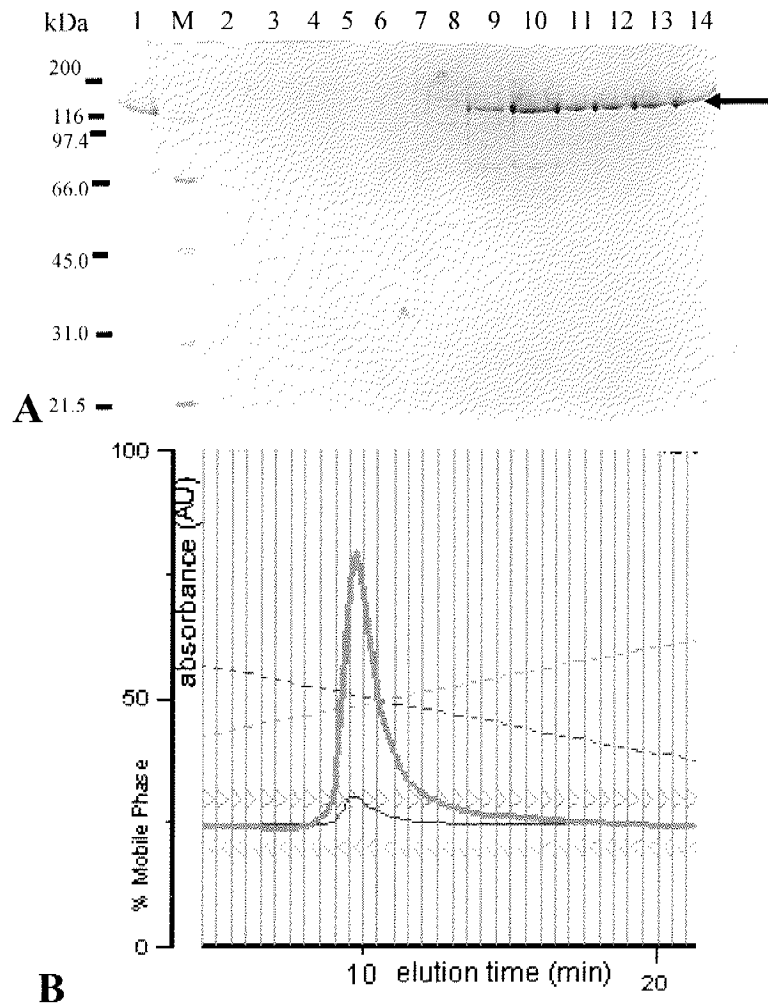


Figure 3.12. hTAF2 Resource S chromatography with TALON™ elution fractions.

A) SDS-PAGE analysis. Lane 1- TALON™ elution fractions (column input), lane 2- flow-through fraction, lanes 3 to 14- peak elution fractions. The arrow indicates the position of hTAF2 band. Lane “M” contains the prestained protein marker (BioRad). B) Chromatogram of the run. The peak contains hTAF2.

3.3.6 Size-exclusion chromatography

Pooled TAF2 fractions, after the Resource S purification, were concentrated to 5 mg/ml using the Amicon device, centrifuged (12000 rpm, 10 min, 4 deg, benchtop) and loaded on the S200 (Superdex 200, 24 ml, Pharmacia) or S6 column (Superose 6, self made, 120 ml, 2.5 cm diameter) at 4° C. In the case of S200 column, only 0.5 ml of the sample was loaded in one run, while up to 2 ml could be loaded on the S6 column. The column was pre-equilibrated in the GF

buffer (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF), or other, depending on the intended usage of the protein. TAF2 was eluted as a single peak, at a position corresponding to the molecular mass of approximately 150 kDa (Figure 3.13 a). If no reducing agent is present from the beginning of the purification, TAF2 will show another peak, roughly corresponding to double of its molecular weight (Figure 3.13 b). This was discovered while using an old bottle of 2-mercaptoethanol for the buffer preparation. The “dimer peak” didn’t disappear completely if the sample was incubated over night in 10 mM DTT, and it was also visible on the SDS gel. Its appearance could be successfully prevented by using fresh 2-mercaptoethanol for the TAF2 lysis buffer and TALON™ buffers. The fact that the higher molecular weight species were never observed suggests that hTAF2 has one accessible cysteine on its surface.

The peak fractions were collected and analyzed by SDS-PAGE. The main 55 kDa contaminant could be removed by the size-exclusion chromatography, as visible on the Figure 3.13 c. Pooled fractions were concentrated to 5-10 mg/ml and used immediately for crystallization, or stored at -80° C flash-frozen in 20% glycerol at the concentration of 3-5 mg/ml.

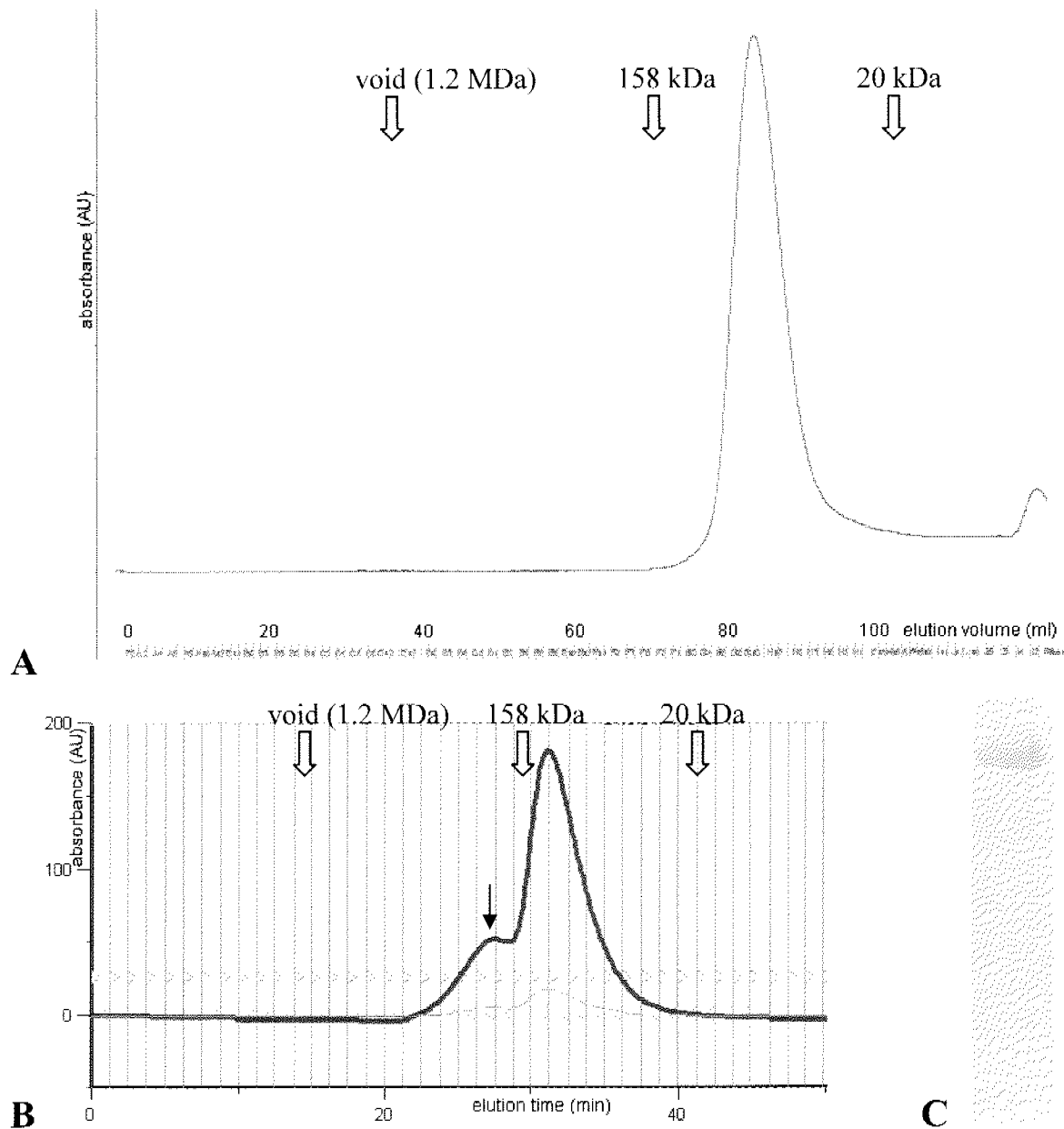


Figure 3.13. hTAF2 size-exclusion chromatography.

A) Run with a normal concentration of the reducing agent (Superose 6 column). B) Run with a low concentration of the reducing agent (Superdex 200 column). The arrow points to the extra “dimer” peak. C) SDS-PAGE of hTAF2 from the peak fractions (monomer).

3.3.7 Analysis of hTAF2 by analytical ultracentrifugation

Analytical ultracentrifugation measurement was done to check if purified hTAF2 is homogeneous and monomeric. Figure 3.14 shows the results of the sedimentation velocity experiment, carried out on the Beckman Optima XL-I. The data was analyzed by numerical

fitting using the program Sedfit. hTAF2 was shown to be mainly monomeric, with the size estimated to be 128 kDa.

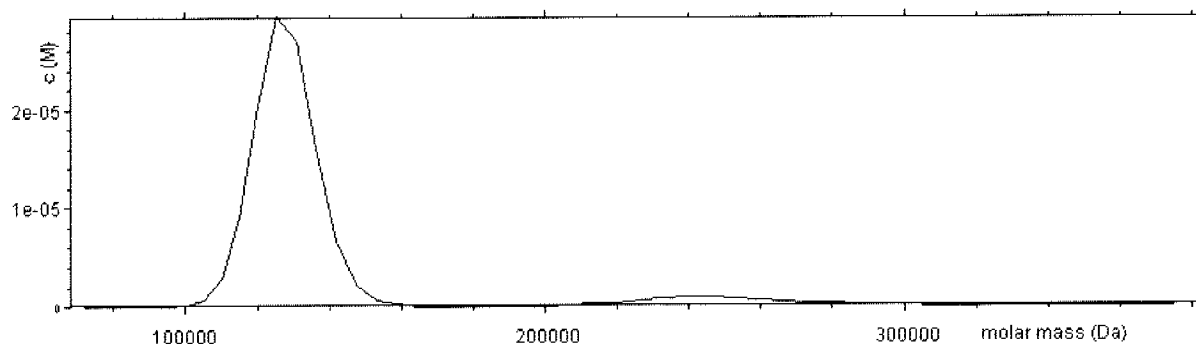


Figure 3.14. hTAF2 analysis by ultracentrifugation (Sedfit analysis).

The main part of purified hTAF2 runs as a monomer. The differential sedimentation coefficient distribution $c(M)$ is plotted against the molecular mass (Da).

3.3.8 hTAF2 purification overview

Figure 3.15 shows an overview of the full-length hTAF2 purification. TAF2 was expressed in Sf21 insect cells and located in the cytosolic fraction. The nuclei were removed as a first purification step, followed by a preparative ultracentrifugation and batch anion-exchange Q-sepharose purification with the soluble fraction. The unbound Q-sepharose fraction, containing TAF2, was loaded on the TALON™ column in the presence of 50 mM imidazole and eluted with a gradient from 50 to 200 mM imidazole. TALON™ elution fractions were used for the cation-exchange Resource S column purification. TAF2 was eluted in a gradient from 0.1 to 1 M KCl and concentrated for the size exclusion chromatography (using S200 or S6 column). The buffers used for hTAF2 purification are listed in Table 3.2.

The final yield was 1- 2 mg per liter of the Sf21 insect cell culture, infected at 2×10^6 cells/ml density.

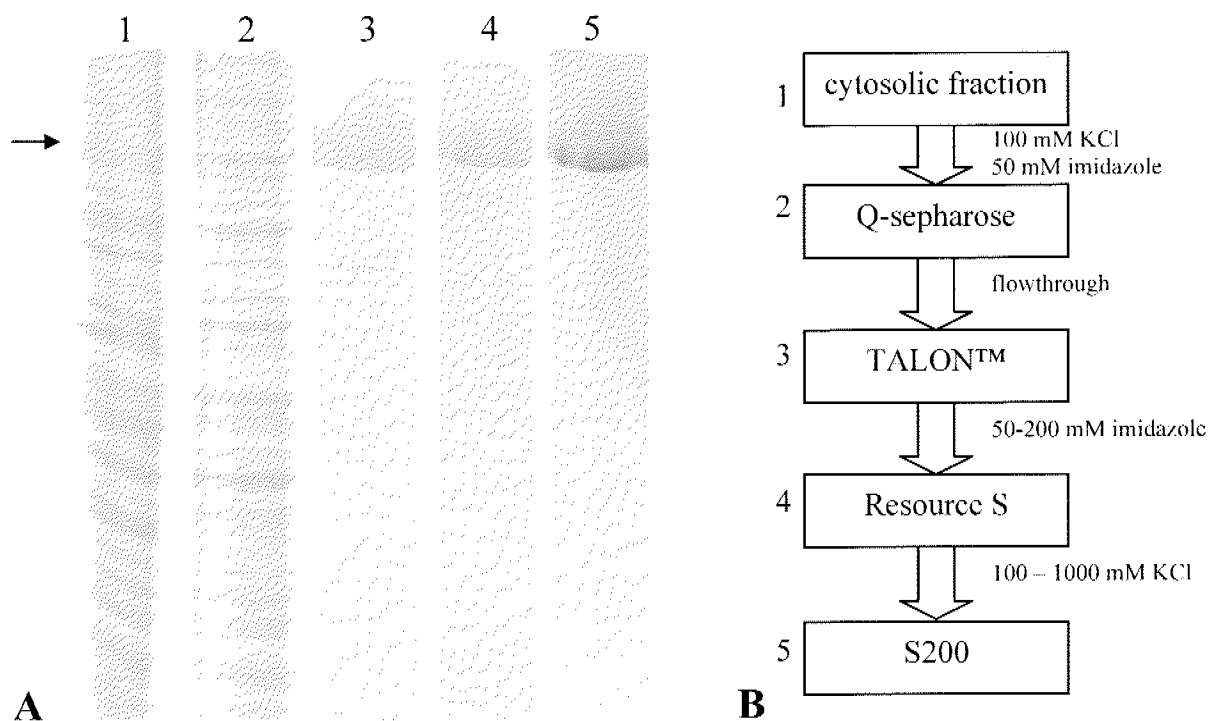


Figure 3.15. hTAF2 purification overview.

A) SDS-PAGE of hTAF2 purification steps (numbers correspond to 5 purification steps, B). The arrow points to the position of hTAF2 band. B) Schematic representation of hTAF2 purification.

BUFFER NAME	COMPOSITION
TAF2 lysis buffer	25 mM HEPES-NaOH pH 7.6 100 mM KCl 10 mM MgCl ₂ 50 mM imidazole-HCl pH 7.6 0.1% NP-40 10 mM 2-mercaptoethanol 1 mM PMSF
TALON™ buffer A	25 mM HEPES-NaOH pH 7.6 100 mM KCl 10 mM MgCl ₂ 10 mM 2-mercaptoethanol 1 mM PMSF
TALON™ buffer B	25 mM HEPES-NaOH pH 7.6 100 mM KCl 10 mM MgCl ₂ 200 mM imidazole-HCl pH 7.6 10 mM 2-mercaptoethanol 1 mM PMSF
TALON™ wash buffer	25 mM HEPES-NaOH pH 7.6 1 M KCl 10 mM MgCl ₂ 10 mM 2-mercaptoethanol 1 mM PMSF
Buffer A	25 mM HEPES-NaOH pH 7.6 100 mM KCl 10 mM MgCl ₂ 1.5 mM DTT 1 mM PMSF
Buffer B	25 mM HEPES-NaOH pH 7.6 1 M KCl 10 mM MgCl ₂ 1.5 mM DTT 1 mM PMSF
GF buffer	25 mM HEPES-NaOH pH 7.6 100 mM KCl 10 mM MgCl ₂ 1.5 mM DTT 1 mM PMSF

Table 3.2. hTAF2 purification buffers

3.4 Discussion

Full-length human TAF2 is a large protein with molecular weight of 137 kDa. The results presented in this chapter show that TAF2 could be successfully produced recombinantly in the insect cell baculovirus system, as well as purified to crystallographic quality from the soluble fraction. The expression of this protein in a prokaryotic system would probably not be possible because of its large size and potential complicated folding. The indication for this is the fact that the proteolytically stable TAF2 core (114 kDa), defined in the next chapter, could not be solubly produced in the *E. coli* system (Appendix II). Therefore, the baculovirus expression system proves again to be the method of choice for large eukaryotic proteins.

The most important step in the full-length hTAF2 purification was the TALON™ affinity chromatography. The stretch of 7 histidines present near the protein's C-terminus proved to be an advantage, since it served as an internal, native His-tag. High concentration of imidazole in the loading buffer was crucial for the purification quality. Without imidazole in the loading buffer, a large number of other proteins from the cytosolic extract would bind nonspecifically to the column and couldn't be completely removed by extensive imidazole washing. The manufacturer's protocol for TALON™ purification suggests adding 5-10 mM imidazole to the loading buffer to prevent non-specific binding, but this concentration was not high enough to ensure high purity of the TAF2 TALON™ elution fractions. In the initial TALON™ purification trials, it was evident that hTAF2 elutes with the higher than usual imidazole concentration (around 150 mM), in contrast to other His-tag fusion proteins which generally elute between 50 and 100 mM imidazole. hTAF2 C-terminal sequence, in addition to the stretch of 7 histidines (instead of 5 in a usual His-tag), contains a histidine-lysine (HK) alternating sequence as well (Figure 3.1), which might explain the strength of binding. Since this sequence should be very soluble, it could be generally used as an alternative to the His-tag, in difficult purifications when the usual His-tag is not effective enough.

Large proteins, such as TAF2, are often proven to be difficult to purify since they are more sensitive to proteolytic degradation and aggregation. TAF2 is a part of the 14-membered multiprotein complex TFIID, and it was likely that it would not be stable and soluble on its own. Some other components of the complex might have been necessary to prevent its aggregation by allowing proper folding or covering hydrophobic patches often present in protein-protein interactions. Nevertheless, TAF2 showed high stability in terms of proteolytic sensitivity (no

degradation products were seen during the purification procedure, which was always done at 4° C within 2-3 days) and solubility (it was soluble in a wide range of buffers, both high and low salt at different pHs). This behavior of TAF2 favors of the possibility that it might not exist exclusively as a part of the TFIID complex, but might function individually as well. It was already suggested that TAF2 is a facultative subunit in the human version of the TFIID complex, since it was shown to dissociate from the complex at high salt conditions of ammonium sulphate precipitation procedure used for TFIID preparation from HeLa nuclear extracts (Martin et al, 1999). Interestingly, this was not seen with the *Drosophila* TFIID. dTAF2 appears to be tightly associated with the other subunits of dTFIID (Verrijzer et al, 1994). Since TAF2 doesn't have a nuclear localization signal, as evidenced by its cytosolic location in Sf21 insect cells and theoretical predictions using PSORT, the question arises of how this protein is transported into the nucleus and assembled into TFIID. It would need an interaction partner to carry it into the nucleus, which opens the interesting possibilities of TAF2-mediated regulation of human TFIID activity and TFIID-unrelated cytosolic TAF2 activity.

4

DEFINITION AND STRUCTURAL CHARACTERIZATION OF hTAF2 DOMAINS

4.1 Introduction

Human TAF2 is a large, multi-domain protein. Since no crystals could be produced of hTAF2 full-length despite extensive trials (Appendix III), it was decided to obtain as much information as possible about its structure using different biochemical and biophysical methods. Results obtained using these methods would provide insights on how to design better crystallization constructs of hTAF2. Since crystal formation is often obstructed due to flexible, disordered regions in the protein structure, like linkers between domains or large loops, it is important to get some information on where the borders between structurally compact, globular pieces are. These structurally (and possibly functionally) independent domains can be defined using both theoretical and experimental methods which supplement each other. Once defined, domains can be produced and manipulated separately, and might prove to be easier to handle than the full-length protein. For example, some of them might be possible to produce and purify from the bacterial expression system, which is simpler and much less time consuming than the insect cell baculovirus system.

One of the most widely used methods to probe a protein's structure is limited proteolysis. This method takes advantage of the fact that exposed protein regions are more likely to be digested by a protease than well folded and compact parts of the protein. The goal of the limited proteolysis analysis using very low protease concentrations that do not digest the whole protein, is to find proteolytically "stable" protein fragments that correspond to structural (and possibly) functional protein domains. This chapter describes how hTAF2 domain structure was defined using limited proteolysis analysis supplemented by other methods.

4.2 hTAF2 domain definition

4.2.1 Analysis of hTAF2 primary structure by bioinformatics

To identify regions in hTAF2 protein conserved through evolution, a BLAST search was performed against the non-redundant protein sequence database (BLAST at NCBI) using the blastp (protein-protein BLAST) algorithm to find TAF2 sequences from different organisms known so far. Table 4.1 shows the resulting sequences as well as the identity percentage (identical amino acids in a sequence alignment) to the human sequence. These sequences were used to obtain a series of multiple sequence alignments which point to the conserved regions and residues in TAF2. Alignment of the human, *Drosophila*, *C. elegans* and rat sequences is shown in the Figure 4.1. Other sequences are omitted for clarity, since loops and gaps found in the less conserved regions from far-away species (especially plants and fungi) can disrupt the alignment. Conserved regions, revealed by this alignment, are found throughout the sequence, except the first 20 – 30 amino acids at the N-terminus and the last 200 amino acids at the C-terminus.

Genbank identifier	Organism	% identity to human TAF2
gi 4507347	<i>H. sapiens</i>	100
gi 124486849	<i>M. musculus</i>	95
gi 71895685	<i>G. gallus</i>	95
gi 76779336	<i>R. norvegicus</i>	91
gi 116256083	<i>D. rerio</i>	85
gi 27924002	<i>D. melanogaster</i>	45
gi 17542624	<i>C. elegans</i>	35
gi 19173129	<i>E. cuniculi</i>	23
gi 25406308	<i>A. thaliana</i>	20
gi 19114243	<i>S. pombe</i>	19
gi 136453	<i>S. cerevisiae</i>	16

Table 4.1. Known TAF2 sequences from different organisms.

The identity percentage to the human sequence is calculated from the quick alignment with BLOSUM matrix (default settings) using DNAMAN.

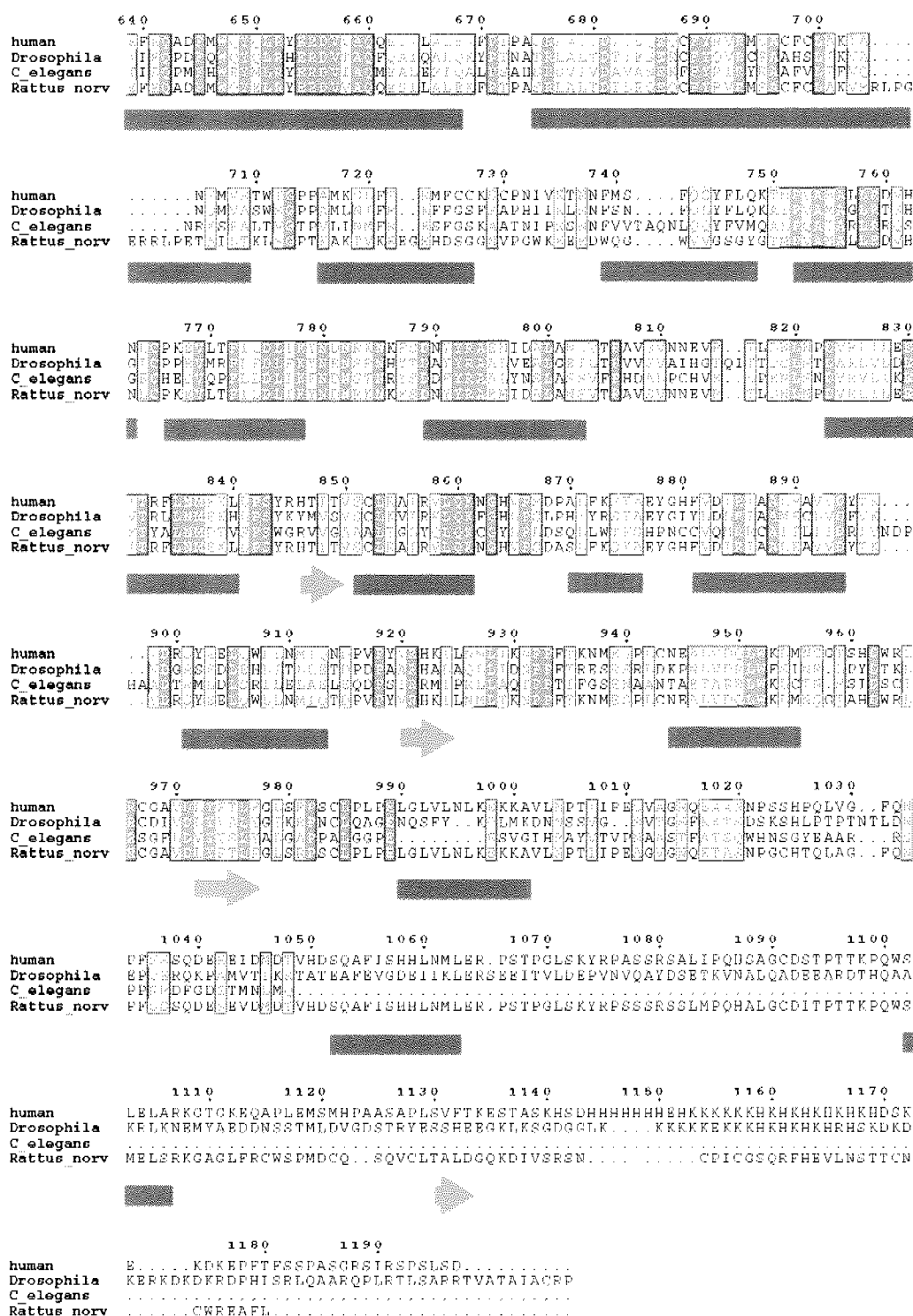


Figure 4.1. TAF2 multiple sequence alignment with secondary structure prediction. Secondary structure prediction was done using the HNN method (Guermeur Y. PhD thesis). Helices are represented as blue bars and beta strands as red arrows.

In addition to the BLAST search, hTAF2 sequence was used for the conserved domain database (CCD) search to identify similar sequences for which 3D structural information is available. The CCD search at NCBI (Marchler-Bauer et al, 2005) found 3 sequences: aminopeptidase N (PepN), puromycin-sensitive aminopeptidase (KOG1046) and aminopeptidase M1 (Figure 4.2). Only the last sequence (aminopeptidase M1) could be completely aligned to the TAF2 sequence (94.1%), while PepN and KOG1046 were 48% and 43% aligned, respectively. Since TFIID or TAF2 have no known aminopeptidase activity, the significance of these results was not clear. Aminopeptidase activity assay was performed in the course of this work, but the results were negative. The CCD search also revealed the region of low complexity at the C-terminal, non-conserved part of the protein. This region corresponds to the series of histidines used for hTAF2 purification (Chapter 3), followed by a series of lysines and an alternate HK sequence that make a cluster of positive charge.



Figure 4.2. Results of the conserved domain database search with hTAF2 sequence. The resulting sequences/structures are shown in red and detected low complexity regions in light blue.

To get more insight into the TAF2 sequence and how the conserved regions could relate to its structure, the secondary structure prediction was performed using several available algorithms. Secondary structure prediction obtained with the HNN (hierarchical neural network) method (Guermeur Y. PhD thesis) is shown in the Figure 4.1, below the multiple sequence alignment. It shows that secondary structure elements are found throughout the sequence, indicating a compact 3D structure. The only exception was the C-terminal non-conserved region of about 200 amino acids which was poor in the secondary structure content and a large 40-residue loop between amino acid 580 and 620. The overall secondary structure content, according to the HNN method, was 38.85% alpha helix, 14.20% extended strand and 46.95% random coil. If the C-terminal 200 amino acids are analyzed separately, they are found to have 86.43% random coil content.

Another method for predicting disordered regions within a protein is DisEMBL (Linding et al, 2003). This method uses three different definitions of disorder: coils/loops as the regions with no secondary structure (but not necessarily disordered), hot loops, a subset of coils/loops with high B-factors (high degree of mobility), and missing coordinates from X-ray structures (defined by Remark-465 entries in Protein Data Bank). According to DisEMBL, there could be several regions of disorder in hTAF2: the N-terminus, region around residue 600, region around residue 800, and the C-terminal region of 100 – 200 amino acids (depending on the predictor).

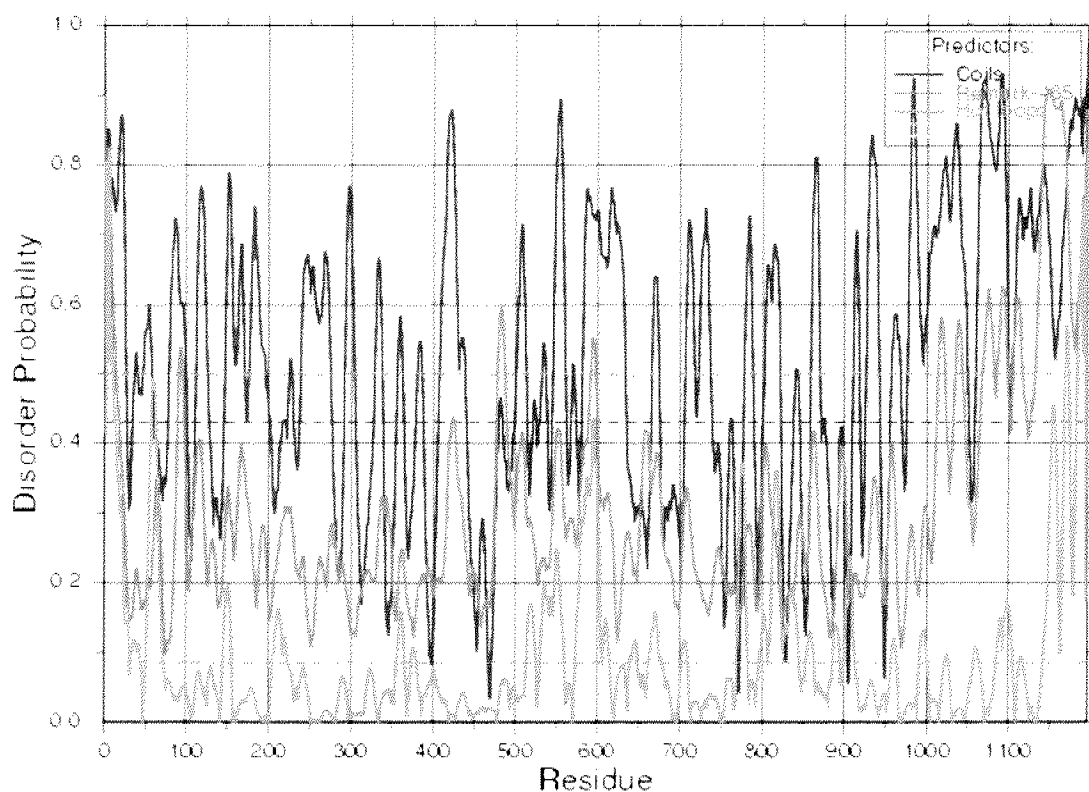


Figure 4.3. hTAF2 disorder prediction with DisEMBL. Disorder probability, according to three different predictors, is shown in different colors. Blue is for coils, green for Remark-465 and red for hot loops. Dotted lines show a random expectation value.

4.2.2 Analysis of full-length hTAF2 by CD spectroscopy

4.2.2.1 Melting temperature determination

Melting temperature (T_m) of hTAF2 was measured by circular dichroism (CD) spectroscopy. Purified hTAF2 at the concentration of 0.2 mg/ml was prepared in the buffer containing 25 mM HEPES-NaOH pH 7.6 and 100 mM KCl. The dependence of the CD signal on temperature was measured and results evaluated as described in the section 2.7.1. Figure 4.4 shows the obtained data and its tentative evaluation performed according to van't Hoff and two-state transition model. The T_m value, determined from the data fit, is 57.4° C. The qualitative shape of the curve fitted shows very cooperative denaturation, indicating initially compact and well folded structure. Since only a single transition is observed, this data indicates either the mutual stabilization or similar stability of hTAF2 subdomains. Since the protein precipitated during heating and could not be reversibly unfolded, the T_m doesn't directly reflect conformational stability but it is influenced also by kinetics of aggregation and the solubility of the unfolded state. Therefore, the enthalpy change for the unfolding could not be determined.

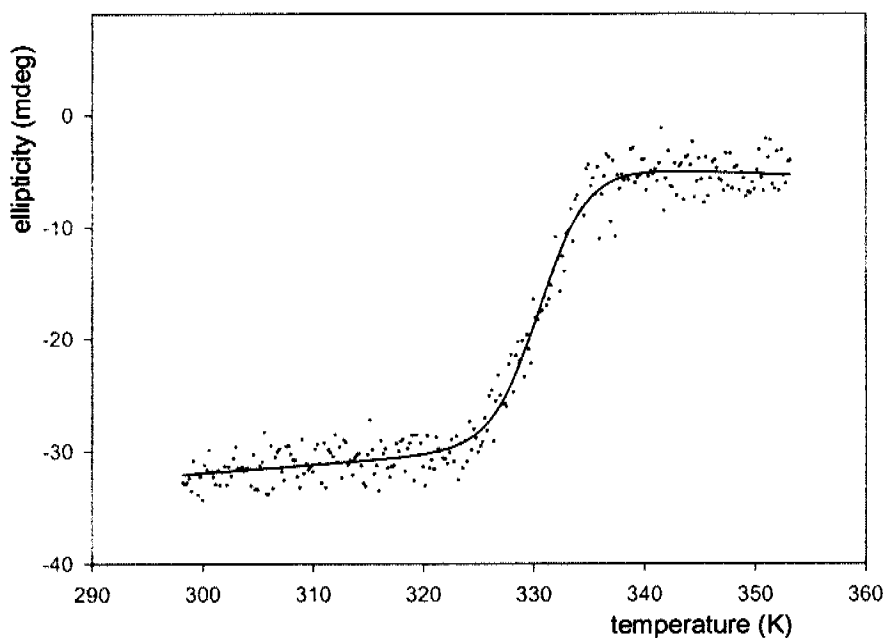


Figure 4.4. Thermal denaturation curve of hTAF2 at pH 7.6.

The solid line corresponds to a tentative fit according to a two-state transition and yielded a T_m of 57.4° C.

4.2.2.2 Analysis of secondary structure content

Circular dichroism (CD) spectroscopy analysis was performed as described in the section 2.7.2 in order to measure the secondary structure content of hTAF2. The full-length protein is predicted to be well structured according to various bioinformatic methods, with the exception of the last 200 amino acids that are predicted to be mainly disordered. Purified hTAF2 sample at a concentration of 0.05 mg/ml was prepared for the CD analysis by an overnight dialysis into CD buffer (10 mM Tris-Cl pH 8.0, 100 mM KF, 2 mM MgCl₂ and 2 mM 2-mercaptoethanol), with one buffer exchange after 1 h. Figure 4.5 shows the final results of the measurement (CD signal with buffer signal subtraction converted to the units of molar ellipticity). hTAF2 clearly shows two minima at 208 and 222 nm, typical for alpha helical structures. The observed signal is weaker than expected ($\sim -4000 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ and $\sim -6000 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$, respectively), since hTAF2 was predicted to have almost 40% alpha helical content (section 4.2.1). Therefore, hTAF2 might contain less than 40% alpha helices.

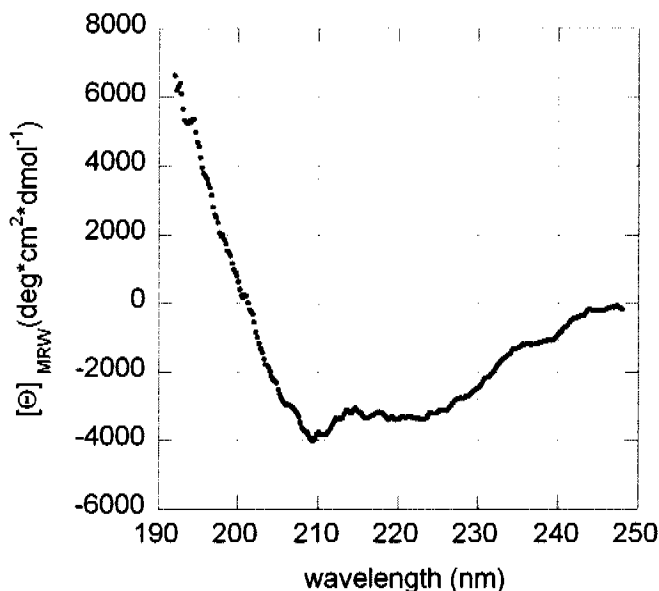


Figure 4.5. hTAF2 secondary structure analysis by circular dichroism. The two minima characteristic for the alpha-helical structures are visible at 209 and 222 nm.

4.2.3 Limited proteolysis of full-length hTAF2

Limited proteolysis analysis on hTAF2 was performed using 3 proteases with different sequence recognition properties: trypsin (cuts after positively charged amino acids), elastase (cuts after small non polar amino acids) and subtilisin (cuts after bulky hydrophobic amino acids). The distribution of the potential cutting sites with these proteases is shown in Figure 4.6.

```

DYKDDDDKN RKKGDKG ES R YK THQV VC NN N QR
KSVVG VE T TVAN NR K NSKQCR YRVR ND EA
A YND T E VCHSESKQRN NY SNAYAA AVSAVD DAG
NGE C KVS E KHVDE K V K H N S DQ KGG H V
V SVEGS AE RGAHV SCGY QNSTR C VDSYSE CT
K E TVDAA VAVSNGD VE TVYTHD RKK T HY T T
AASN S A G E VD Y HEVTH C Q KHTTS
Y HEV E YE E TCRY YS C KTV DEA YVEVAAYAS
S STN HS A DET T RRC AQS AQ Q GC SR
S SDE V KG SGY YG KKT GVNEYR H KEE DK
VAYE KTGGV H GGGK EKDN ASH H S KH HT S
EYTT QCK AH V R EN R S E QV NK S AST
ASSQK QSH SQ VSTSG K S SNVSG KD Q KQ
VDQSGVVK Y GS A NRKRN V E E KQDY TS GTQKYVG
KVTVQE D GS NHT Q E ENS KHD C HSKSRRNKKK
K NGEEV D D SA DAD S R D D SV RKVE
EQAD QYQ RYERDVVAQ QES EA EK T ASR A T
D EQEQC Y RVR SAC C AK ANS VST TG AKS
TR CCKSC N VKTNN S QSY QKT VA RD
VHN C KEV T D KYN DNRKNK SDN YYRAE DA
ANSVT AVSV NNEVRT DN N DVR EE TR N EK
SYRHT TV SC RA RV Q KNGHV SD A KSYAEYGH
VD R AA E AVVDYTKVDR SYEE Q N QND V YV
RHK N TK N TKN ES CNEA VDQ K NSGTS
HD R RCGAV D Y T G S R SC E G V N KEKK
AV N T E SVAGNQEAAN N SSH Q VG QN SSSQD
EEE D DTVH DSQA SHH N ER ST G SKYR ASSR
SA QHSAG CDST TTK Q S E ARKGT GKEQA E S
H AASA S V TKESTASK HSDHHHHHHH EHKKKKKKHK
HKKKKKKKHD SKEKDKE T SS ASGRS RS S SD

```

Figure 4.6. Distribution of protease cutting sites in the hTAF2 sequence.

Red – trypsin (RK), blue – elastase (AVG), yellow – subtilisin (ILMFPW). The figure is made with The Protein Colourer at <http://www.ebi.ac.uk/cgi-bin/proteincol/ProteinColourer.pl>.

Proteolytic digests were performed with purified hTAF2 as described in section 2.6.6. A series of tests with different protease:protein mass ratios were performed in order to obtain the protease concentration that can produce the most informative results (a wide range of fragments covering different molecular masses). For the tryptic digest, 1:1000 (protease:protein) mass ratio was used and the samples were taken 15, 30, 60 and 120 min after being exposed to the protease. Figure 4.7 shows the results of the tryptic digest analyzed by 12% SDS-PAGE. At least 9 protein

fragments could be distinguished, some of them being more pronounced and stable over time than others: TR-2, TR-3 and TR-6. These proteolytically stable fragments, with molecular masses of about 110, 90 and 20 kDa (determined from the SDS-PAGE), respectively, were identified as potential hTAF2 domains.

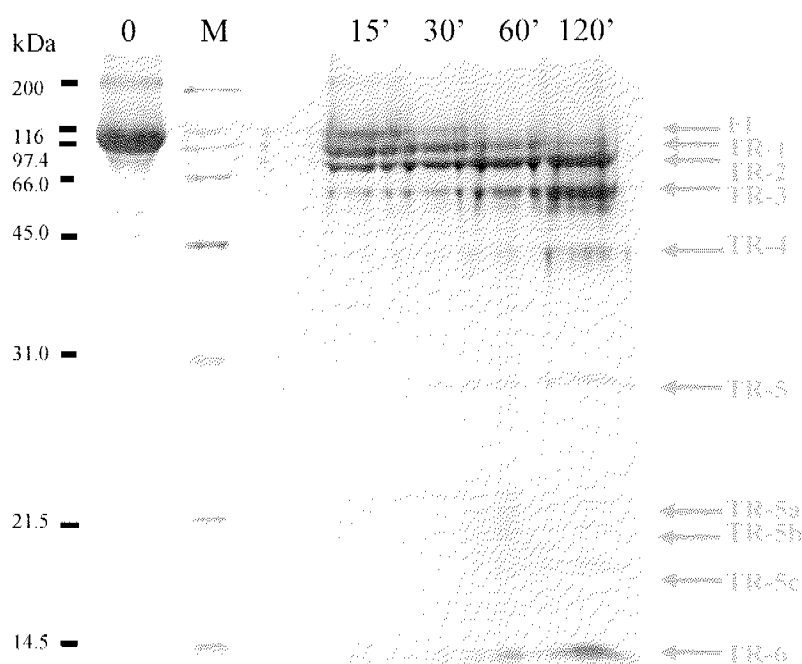


Figure 4.7. Time course of hTAF2 tryptic digest.

The sample before trypsin addition is marked by “0”. M corresponds to the BioRad Broad Range protein marker. The arrows show the position of proteolytic fragments, as well as the full-length protein (FI).

To confirm the results obtained by the tryptic digest of TAF2, the digest with elastase (mass ratio 1:2000) and subtilisin (mass ratio 1:10000) was performed. In the case of subtilisin, it was necessary to perform several digests with different protease concentrations since the reaction equilibrates rapidly and therefore the time course doesn't reveal any new fragments. The results obtained with both proteases showed two stable fragments of 110 and 90 kDa, exactly as for the tryptic digest (Figure 4.8).

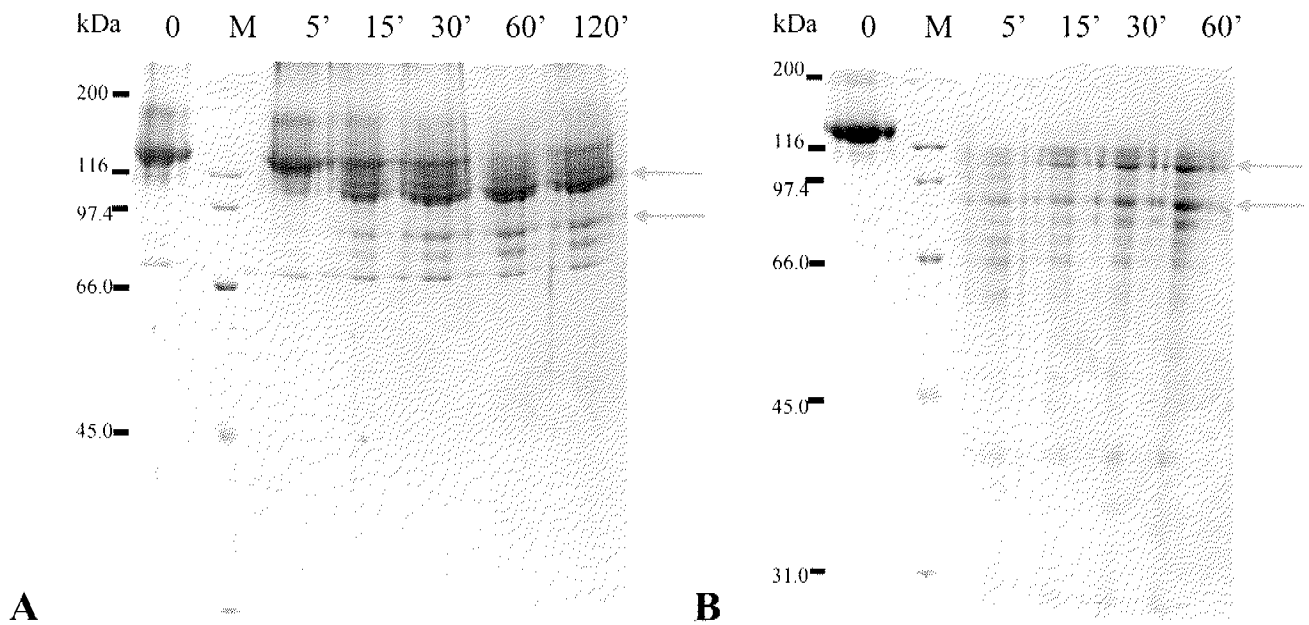


Figure 4.8. Time course of hTAF2 digest with elastase (A) and subtilisin (B). The sample before protease addition is marked by “0”. M corresponds to the BioRad Broad Range protein marker. The arrows point to the 110 and 90 kDa proteolytic fragments.

Detailed analysis of tryptic fragments was performed in order to identify them and get the information on the location of disordered regions. The method used for fragment identification was N-terminal sequencing of SDS-PAGE separated fragments blotted on the PVDF membrane. Molecular weight was estimated from the SDS-PAGE to get information on the fragments' C-termini. Several challenges were met in this process, making some fragments very difficult to identify. One problem comes from the fact that proteins produced in Sf21 insect cells have an N-terminal block due to post-translational modification, making them impossible to sequence by Edman degradation. To circumvent this problem, tryptic fragments of hTAF2 were also analyzed by Western blot using the anti-Flag antibody (hTAF2 construct used in this work had an N-terminal Flag-tag). This method identified several proteolytic fragments coming from the N-terminus from TAF2: TR-1, TR-2, TR-3 and TR-4 (Figure 4.9). No other fragments, in the lower molecular weight range, were identified to have retained the N-terminal Flag-tag. Another technical difficulty came from the fact that the bands from the high molecular weight range are less efficiently blotted on the PVDF membrane. Because of their size, there is also more chance these bands would contain a mixture of different fragments making them technically difficult to sequence. Table 4.2 summarizes the information obtained on hTAF2 tryptic fragments.

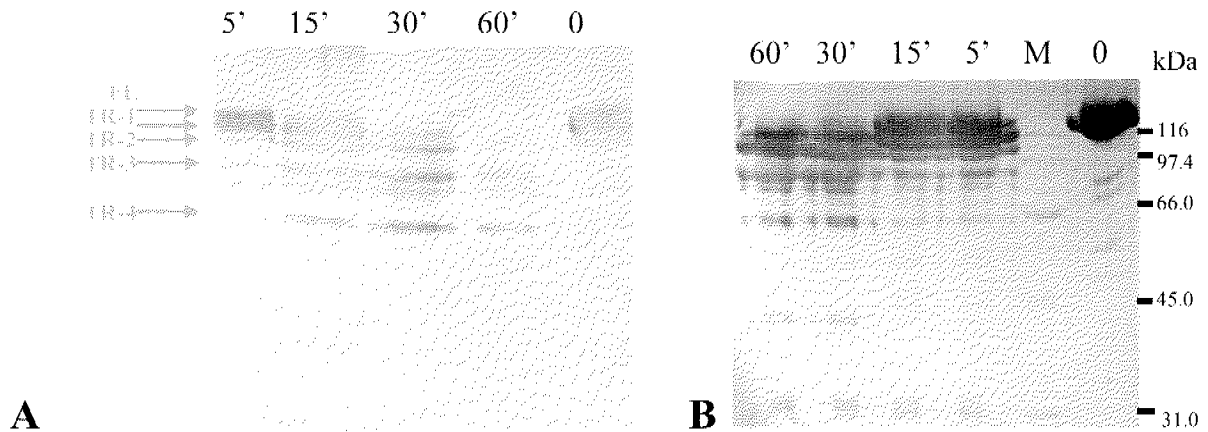


Figure 4.9. Identification of the N-terminal tryptic fragments by anti-Flag Western blot. A) PVDF membrane of the Western blot. The arrows show the fragment positions. B) SDS-PAGE of the same samples.

Fragment	Size (kDa)	N-terminal sequence	Start (aa)	End (aa)	Sequencing quality
FL	137	-	1	1197	-
TR-1	~120	-	1	?	-
TR-2	~110	rKKGD	1 or 12	~1000	bad
TR-3	~90	-	1	816?	-
TR-4	~60	rKKGD	1 or 12	597?	bad
TR-5	~40	rNKKKK	597	~1000	good
TR-5a	~35	rTLDNL	816	?	good
TR-5b	~30	rTLDNL	816	?	good
TR-5c	~25	kTNNFM	736	~1000	good
TR-6	~20	rTLDNL	816	~1000	good

Table 4.2. Results of the N-terminal sequencing of TAF2 full-length tryptic fragments. Fragment sizes are estimated by SDS-PAGE. The N-terminal sequence, as obtained by Edman degradation, is shown in capital letters. The final amino acid (“End”) is estimated from the fragment size, position of the potential tryptic cutting sites and the relation to other fragments.

4.2.4 Model of hTAF2 domain structure

Despite the incompleteness of the information on the tryptic fragments of hTAF2 obtained by limited proteolysis analysis, it was possible to make several conclusions about hTAF2 domain structure (Figure 4.10). Overall, hTAF2 appeared to have well folded, compact

domains as indicated by limited proteolysis. From digests with trypsin, elastase and subtilisin, it was clear that the proteolytically most sensitive part are is C-terminal 200 amino acids. This “TAF2 tail domain”, which is first to be degraded, is the region poor in secondary structure elements and contains the internal “His-tag” used for the purification. The fact that this histidine stretch could be used for purification indicates that it is exposed. Other sensitive spots are the N-terminal 11 amino acids (Flag-tag and enterokinase site) and the region around the position 816. The middle region between the putative aminopeptidase domain and the second half of the protein is also relatively sensitive. The most stable fragments appear to be TR-2, TR-3 and TR-6. Since they might correspond to separate TAF2 domains, they were taken as a basis for new crystallization construct design (described in the next section).

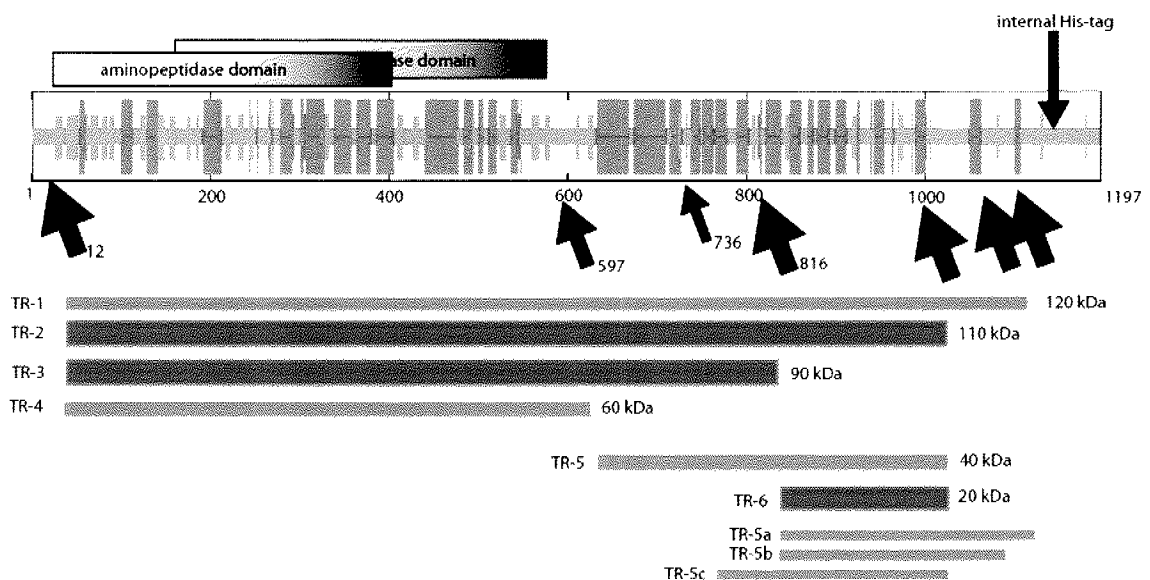


Figure 4.10. Summary of the limited proteolysis results and theoretical structural predictions on hTAF2 sequence.

hTAF2 sequence (amino acid 1-1197) with the secondary structure prediction (α helix – blue, β strand – red, random coil – violet) is represented as a horizontal bar. Tryptic hTAF2 fragments are shown below. Thickness of the fragment represents its apparent stability. Detected tryptic cutting sites are shown as an arrow, with a number in a case N-terminal sequencing revealed the exact cutting position. The size of the arrows indicates the apparent cutting efficiency.

4.3 Production and characterization of the hTAF2 core domain

4.3.1 Design, production and the analysis of hTAF2 “TEV” constructs

In an attempt to obtain a crystallizable piece of hTAF2 that corresponds to its structural core, six constructs were designed based on the limited proteolysis analysis of the full-length protein described in the previous section. There were two possible strategies for production of these constructs: 1) introducing a STOP codon at the desired position along with a new affinity purification tag (TAF2 “tail” with the internal His-tag is a part of the disordered region which would be omitted in these constructs), and 2) insertion of a TEV protease recognition site at the desired position (Figure 4.11), allowing a modified full-length protein to be produced and purified according to the already established purification protocol (Chapter 3). The second strategy was chosen because it is less time consuming. There would be fewer cloning steps and no need to develop a new purification protocol. Also, it would permit production of a soluble protein in a quantity similar to the wild-type hTAF2 protein. The second strategy would make it possible to get all the resulting protein fragments after the TEV cleavage, since the C-terminal piece with the His-stretch could be separated from the rest using TALON™ column purification (Figure 4.12).

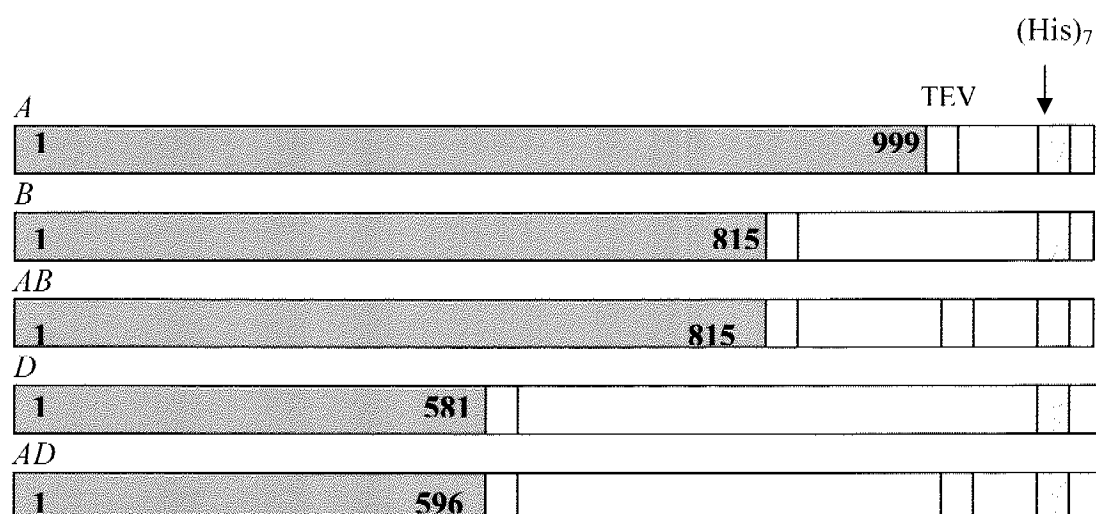


Figure 4.11. hTAF2 TEV constructs (A, B, AB, D, AD).

Inserted TEV protease recognition site is shown as a yellow bar and internal histidine stretch as a pink bar. Start and end positions of the fragment before TEV site is indicated.

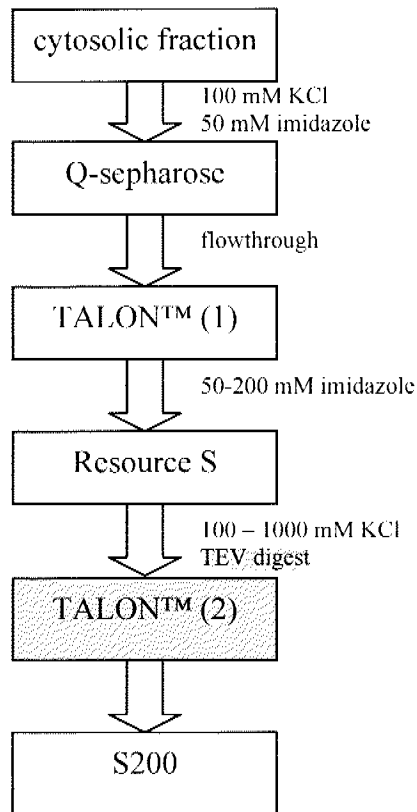


Figure 4.12. TAF2 TEV constructs purification summary.

Steps different from the full-length TAF2 purification are marked yellow. Second TALON™ purification step allows the fragment containing the His-stretch to be separated from the rest.

4.3.1.1 Cloning of hTAF2 “TEV” constructs

The exact positions of the TEV site insertions were chosen to be at amino acid 999 (position “A”, assumed end of the TR-2 fragment), amino acid 815 (position “B”, strong tryptic cutting site), amino acid 581 (position “D”, before the tryptic cutting site but outside of the conserved region) and amino acid 596 (second version of position “D” - “D2”, detected tryptic cutting site). AB and AD constructs contained TEV sites at two positions.

Method chosen for the TEV site introduction was quick change mutagenesis (the adapted protocol is described in the section 2.2.21). Since TEV site requires a very large insertion of 24 bp (Figure 4.13), the primers had to be designed carefully, considering the number of mismatched positions and the melting temperature of a primer-template hybrid. Therefore, they had around 25 residues flanking the TEV sequence on each side (the exact length was suggested by the software available at the time on the Stratagene web pages). Additionally, a *Bam*H I site

was designed within the inserted region to facilitate clone screening. This strategy was used for the production of constructs A, B, D and D2, while the constructs AB and AD were produced by subcloning, combining the single TEV site constructs. Single TEV constructs were digested with *EcoR* I, giving two fragments which were gel purified. The larger fragment from the construct A (containing the TEV site) was dephosphorylated and ligated with the small (TEV site-containing) fragment from the construct B or D, resulting in constructs AB or AD, respectively. The correct insert orientation was checked by *Pvu* II digest and the presence of both TEV sites was confirmed by *Bam*H I. Finally, all 6 constructs were confirmed by sequencing through the full-length of the reading frame.



Figure 4.13. TEV recognition sequence used in the TAF2 constructs. *Bam*H I site, used for clone screening, is underlined.

4.3.1.2 Expression and purification of hTAF2 TEV construct A

TAF2 construct A was expressed in Sf21 insect cells using the same procedure as described for the wild-type full-length TAF2 (Chapter 3), resulting in a comparable expression level (Figure 4.14 A, lane 1). First 4 purification steps (cytosolic extract preparation, Q-sepharose, TALON™ and Resource S chromatography) were done essentially the same as with the wild-type TAF2. The only exception were the Resource S buffers A and B, which contained 2-mercaptoethanol as a reducing agent instead of DTT, since the eluted sample was to be used for TALON™ chromatography. All the chromatographic properties of this construct were exactly the same as the wild-type TAF2 (Figure 4.14). Gel filtration run with the Resource S fractions showed the same results as with the wild-type protein (construct A eluted after 11.5 ml on the S200 column as well).

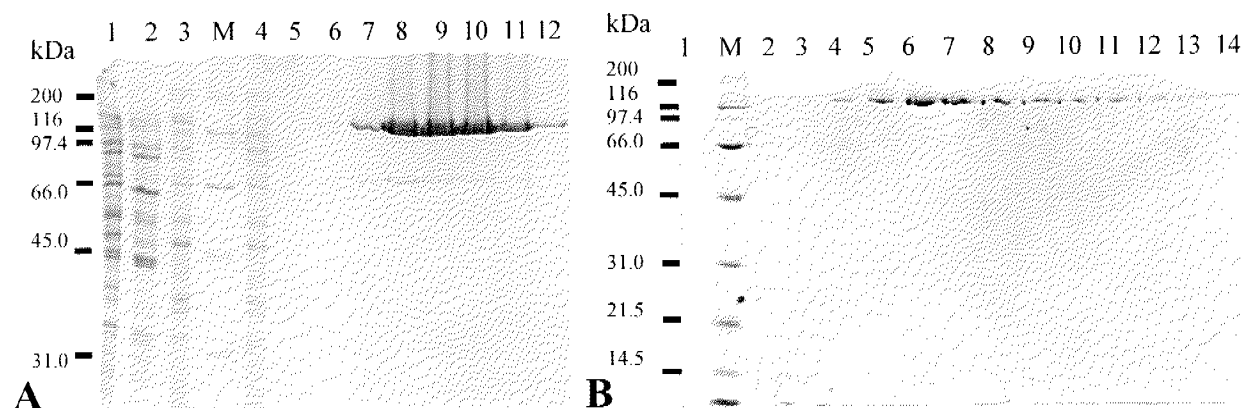


Figure 4.14. TAF2 TEV construct A expression and purification.

A) Expression, Q-sepharose and TALON™ purification. Lane 1- cytosolic extract, 2- Q-sepharose 1 M KCl eluate, 3- Q-sepharose unbound fraction, 4- TALON™ flowthrough, lanes 5-12 TALON™ elution fractions.

B) Resource S chromatography. Lane 1- column input, lanes 2-14 – elution fractions. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Resource S elution fractions were pooled and the protein concentration was determined. TEV protease was added to the mass ratio of protease:protein of 1:10. Samples were taken for analysis after 1, 4 and 12 h to determine the time necessary for the complete digest, which was subsequently always done over night at 4° C (Figure 4.15). The digest yielded two fragments, 116 and 22 kDa in size, as expected. Interestingly, the smaller fragment showed unusual electrophoretic properties, traveling next to the 31 kDa protein marker on the SDS-PAGE. The reason is probably the large number of lysine residues in this region, making a cluster of positive charge that may result in a slower mobility of this fragment towards the positive electrode.

After the TEV digest, another TALON™ purification step was performed in order to separate the two fragments. Fragment mixture, containing the TEV protease as well, was loaded on the TALON™ column pre-equilibrated in the TALON™ buffer A. The large fragment was present in the unbound fraction and the small one was eluted with a linear 0-200 mM imidazole gradient between TALON™ buffer A and B, along with the TEV protease (Figure 4.16). A small amount of the large fragment was bound to the TALON™ column nevertheless, probably complexed with the TEV protease as they co-elute.

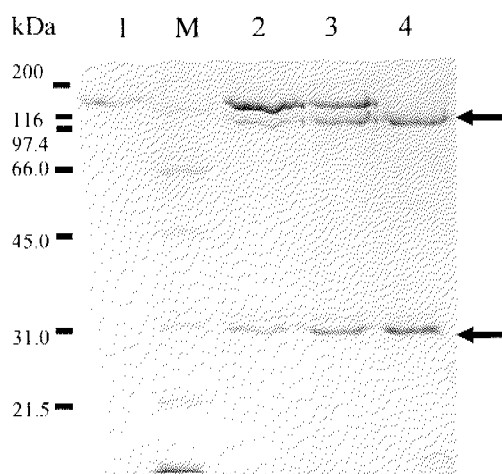


Figure 4.15. TAF2 construct A TEV digest.

Lane 1- sample before TEV addition. Lanes 2, 3, 4- samples taken after 1, 4, 12 h digest at 4° C, respectively. Lane M contains the broad range SDS-PAGE protein marker (BioRad). The arrows point to the positions of two resulting TAF2 fragments. A faint band above the smaller fragment corresponds to the upper band of the TEV protease (the lower one overlaps with the TAF2 small fragment).

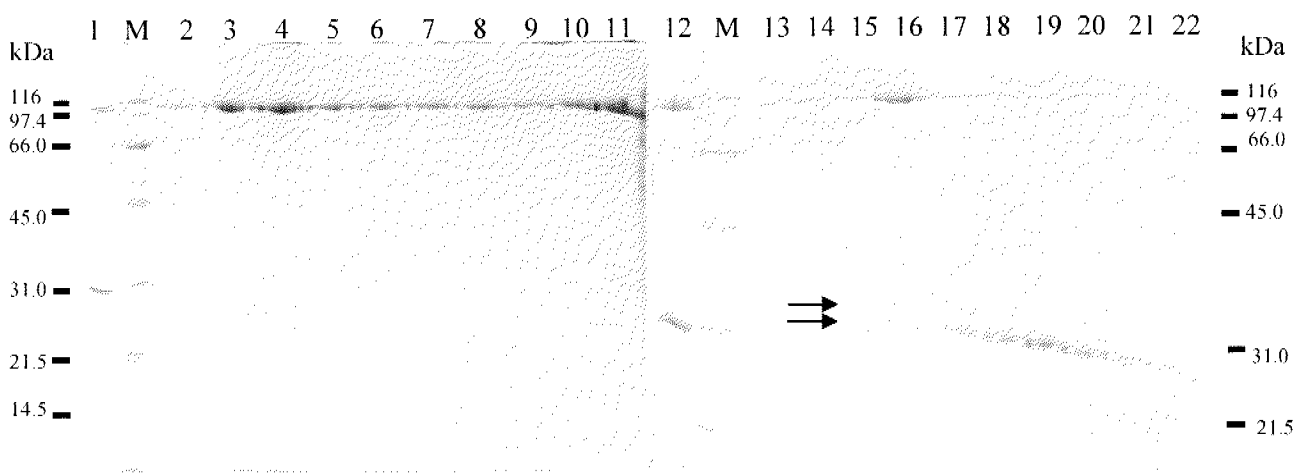


Figure 4.16. Separation of two construct A fragments by TALON™ chromatography.

Lanes 1 and 12- column input, lanes 2 to 11- unbound fractions, lanes 13 to 22- elution fractions.

Arrows point to the TEV protease bands. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

A gel filtration run with the TEV digest mixture was performed as well, using the S200 column pre-equilibrated in the GF buffer (Figure 4.17). Fragments eluted from the column separately, at the position expected for their molecular weight, indicating that they don't interact. The larger fragment (116 kDa) eluted after 12.5 ml and the smaller (22 kDa) after 14 ml while

the full-length hTAF2 eluted after 11.5 ml on the same column. The 22 kDa fragment had to be pooled and gel filtrated again in order to purify it away from the overlapping large fragment peak. The small fragment was concentrated to 3 mg/ml, and the big fragment to 10 mg/ml before the final SDS-PAGE analysis (Figure 4.17 B). The analysis revealed several degradation fragments below the small fragment, as well as the possible disulphide dimer band and small amount of the large fragment.

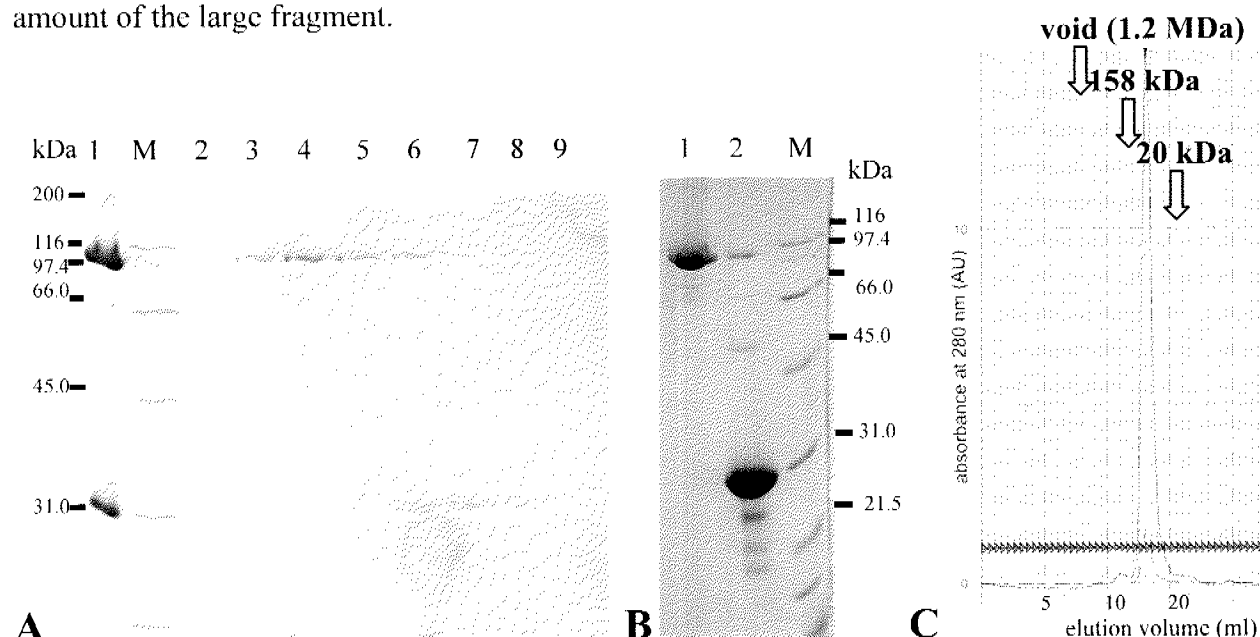


Figure 4.17. Gel filtration with the hTAF2 construct A TEV digest mixture.

A) SDS-PAGE of the elution fractions. Lane 1- column input. Lanes 2 to 9- 0.5 ml elution fractions from 11.5 to 15 ml. B) Pooled and concentrated fractions of two fragments. Lane 1- 116 kDa fragment, lane 2- 22 kDa fragment. Lane M contains the broad range SDS-PAGE protein marker (BioRad). C) Chromatogram showing one main 116 kDa fragment peak with a small shoulder from the 22 kDa fragment. 0.5 ml fractions are represented by horizontal lines.

4.3.1.3 Expression and purification of hTAF2 TEV construct B

TAF2 construct B was expressed in Sf21 insect cells using the same procedure as described for the wild-type, full-length TAF2 (Chapter 3), resulting in a comparable expression level, and first 4 purification steps were done exactly as described for TAF2 construct A (section 4.3.1.2). Like TAF2 construct A, the construct B had the same chromatographic properties in these steps as the wild-type TAF2 (Figure 4.18). Gel filtration of the Resource S fractions also showed the same results as with the wild-type protein and construct A, since construct B eluted after 11.5 ml on the S200 column.



Figure 4.18. TAF2 TEV construct B expression and purification.

A) Q-sepharose and TALON™ purification. Lane 1- Q-sepharose unbound fraction, 2- TALON™ flowthrough, lanes 3-13 TALON™ elution fractions. The arrow points to the position of the expressed protein band in the Q-sepharose unbound fraction. B) Resource S chromatography. Lane 1- column input, lanes 2-14 – elution fractions. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Resource S elution fractions were pooled and the protein concentration was determined. TEV protease was added to the protease:protein mass ratio of 1:10 and the sample was incubated overnight at 4° C. Interestingly, the complete digest could not be achieved under these conditions which worked well for construct A. This result indicated that the TEV site at the position B is less accessible to TEV protease. The whole sample was digested only after 4 days at 4° C, after adding twice the amount of TEV protease. The digest yielded two fragments, 94 and 44 kDa in size, as expected (Figure 4.19, lane 1).

Figure 4.19 shows the results of the TALON™ chromatography with the TEV digest mixture, done exactly as with the TEV digest mixture of construct A (section 4.3.1.2). Surprisingly, the two fragments could not be separated, as the 94 kDa fragment (no His-tag) co-eluted with the smaller fragment (carrying an internal His-tag). The fragments eluted exactly at the same position as the non-digested protein, at 150 mM imidazole concentration. To confirm the results obtained with the TALON™ chromatography, construct B TEV digest mixture was loaded on the S200 gel filtration column pre-equilibrated in the GF buffer (Figure 4.20). The fragments co-eluted again exactly at the position (11.5 ml) of the non-digested protein.

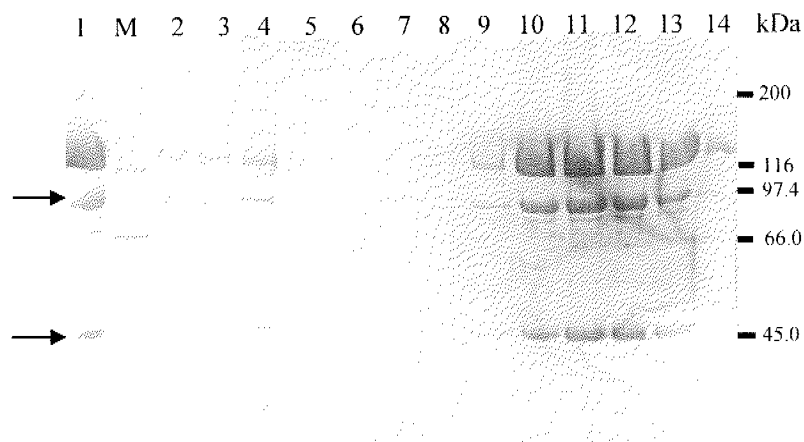


Figure 4.19. Separation attempt of two construct B TEV digest fragments by TALON™ chromatography. Lane 1- column input, lanes 2 to 5- unbound fractions, lanes 6 to 14- elution fractions. Arrows point to the 94 and 44 kDa fragment bands. Lane M contains the broad range SDS-PAGE protein marker (BioRad). Small amount of sample is found in the unbound fractions since the column capacity was exceeded.

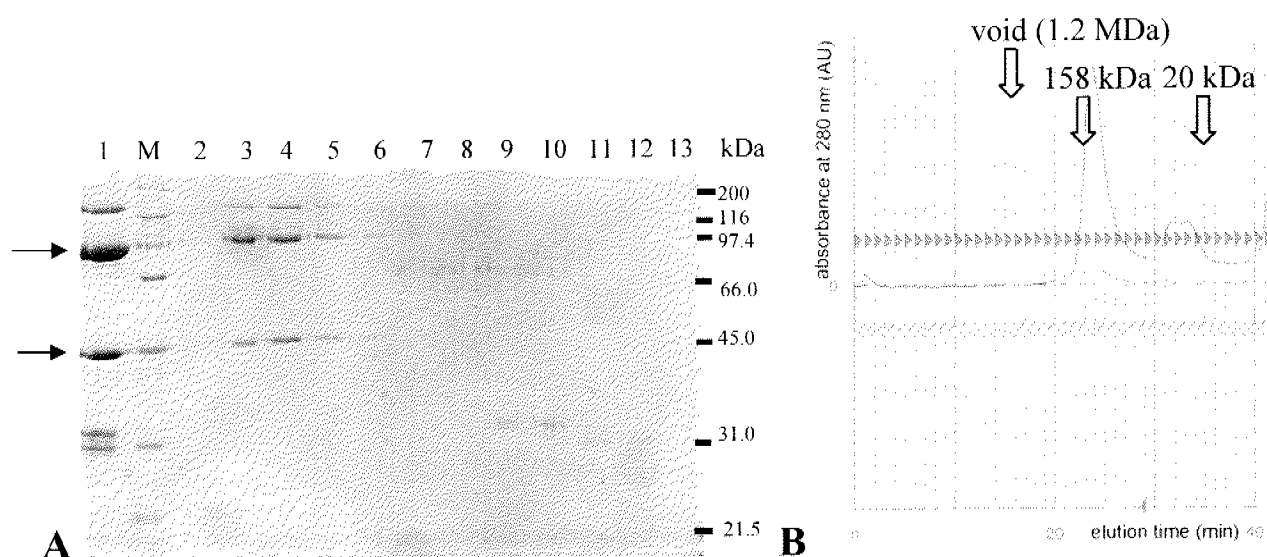


Figure 4.20. Gel filtration with the hTAF2 construct B TEV digest mixture.

A) SDS-PAGE of the elution fractions. Lane 1- column input. Lanes 2 to 13- 0.5 ml elution fractions from 10.5 to 16 ml. Lane M contains the broad range SDS-PAGE protein marker (BioRad). The arrows point to the 94 and 44 kDa fragments, resulting from the TEV digest. The highest band is the undigested protein and the lower double band corresponds to the TEV protease. B) Chromatogram shows the main peak with the undigested and digested construct B fragments, as well as the smaller double peak of the TEV protease.

In an attempt to separate the fragments, they were loaded on the TALON™ column again and washed with several buffers which contained up to 2 M KCl and 1 M urea. None of these procedures was successful in separating the fragments which co-eluted with the 150 mM

imidazole from the TALON™ column. These results indicate that these two fragments interact strongly and are probably co-folded. Therefore, the 94 kDa fragment does not constitute a structurally separate domain of hTAF2 protein.

4.3.1.4 Expression and purification of hTAF2 TEV constructs AB, D, AD and D2

TAF2 TEV constructs AB, D, AD and D2 were expressed in Sf21 insect cells resulting in an expression level comparable with the wild-type TAF2. All constructs behaved the same with the first 4 purification steps, which were done exactly as described for TAF2 construct A (section 4.3.1.2). Their chromatographic properties in these steps were the same as the wild-type TAF2, including the gel filtration results with the Resource S fractions.

TAF2 construct AB was purified in large scale for an attempt of crystallization (Appendix III). Its expression and purification, including the TEV digest, is shown in Figure 4.21. This construct contains two TEV sites, combining constructs A and B. The complete TEV digest produces 3 fragments, one of 94 kDa (the same as the big fragment from construct B) and two 22 kDa fragments (between TEV sites B and A and the C-terminal tail fragment). AB construct confirmed the results already obtained with the construct A and B: the C-terminal 22 kDa fragment, containing the internal His-tag, could be separated from the rest using the TALON™ chromatography, while the other 22 kDa fragment strongly interacted with the 94 kDa fragment.

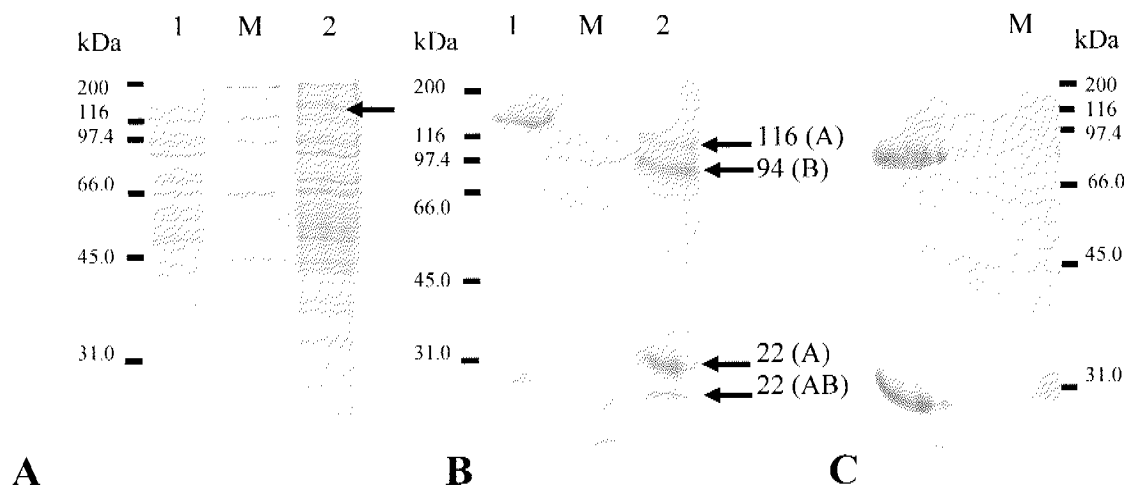


Figure 4.21. TAF2 construct AB expression and purification.

A) Expression in Sf21 insect cells. Lane 1- non-infected cells, 2- infected cells. The arrow points to the expressed AB construct band. B) Purification of the AB construct. Lane 1- pooled Resource S fractions, 2- TEV digested sample with arrows pointing to the fragments resulting from single or double cuts. The upper 22 kDa band comes from the C-terminal highly positively charged fragment and travels more slowly in the gel. C) Final purified sample containing the 94 and 22 kDa fragments without the 22 kDa C-terminal His-tag containing fragment. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

TAF2 constructs D and AD were made in order to find out if the first 600 amino acids in TAF2 constituted a separate domain, as expected from the bioinformatics analysis. The TEV digest of these purified constructs would allow obtaining several TAF2 fragments interesting for structural analysis. Test expression and batch purification were made with both constructs in parallel (Figure 4.22). First purification step was done with the TALON™ resin and TALON™ elution fractions were used for the SP-sepharose batch purification. The proteins eluted from the SP-sepharose with the buffer containing 400 mM KCl and subsequently used for the TEV digest. Unfortunately, the TEV protease wasn't able to cut at the position D, indicating that this TEV site is inaccessible. The A position within the AD construct was cut normally, acting as an internal positive control of the TEV protease activity (Figure 4.22 C).

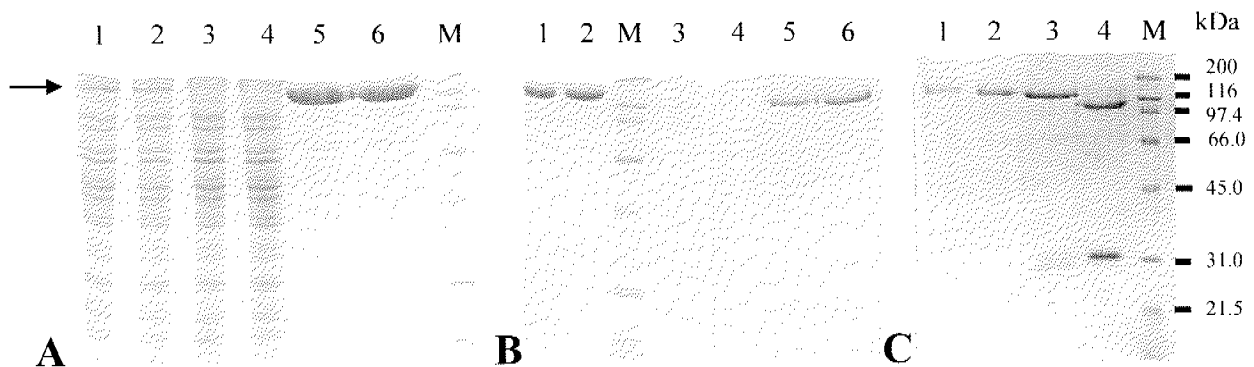


Figure 4.22. TAF2 “TEV” constructs D and AD test expression and purification.

A) Expression and TALON™ batch purification. Lanes 1, 3, 5- cytosolic extract, unbound TALON™ fraction and TALON™ elution fraction of construct D, respectively. Lanes 2, 4, 6- cytosolic extract, unbound TALON™ fraction and TALON™ elution fraction of construct AD, respectively. The arrow points to the expressed band position. B) SP-sepharose batch purification. Lanes 1, 3 and 5- input, flowthrough and elution fractions of the construct D, respectively. Lanes 2, 4 and 6- input, flowthrough and elution fractions of the construct AD, respectively. C) Construct D and AD test TEV digest. Lanes 1 and 3- construct D before and after TEV incubation, respectively. Lanes 2 and 4- construct AD before and after TEV incubation, respectively. TEV protease can cut at the position A, but not D. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Construct D2 was designed in another attempt to obtain a protein with an accessible TEV site at the position D. For this reason, the TEV site was inserted 15 amino acids after the initial D site, exactly at the position of the detected tryptic cutting site. This construct was purified over the TALON™ column and incubated with the TEV protease. Like in the case of the construct D, the protease wasn't able to cut this construct, indicating that the TEV site is buried and inaccessible (Figure 4.23).

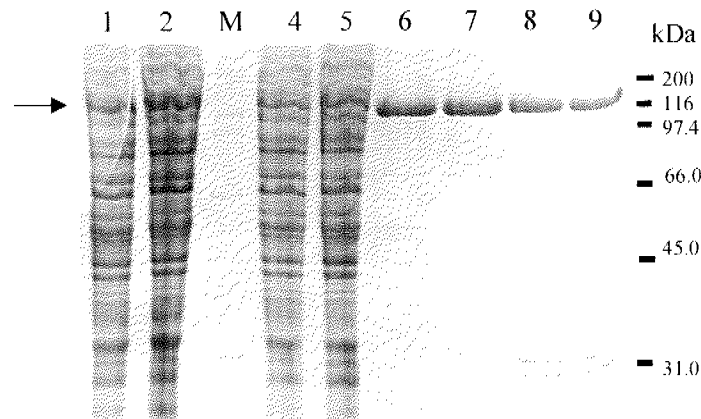


Figure 4.23. TAF2 “TEV” construct D2 test expression and purification.

Lanes 1 and 2, 4 and 5, 6 and 7, 8 and 9 show the expression, TALON™ flowthrough fractions, TALON™ elution fractions and TEV digests, respectively, from two different D2 construct viruses. The arrow points to the position of the expressed protein band. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

4.3.2 Analysis of the hTAF2 TEV construct A

4.3.2.1 Limited proteolysis experiments

Based on the conclusions made from the hTAF2 TEV construct analysis, a working model was made that hTAF2 construct A 116 kDa fragment was the hTAF2 core domain. It was decided to proceed with crystallization and further biochemical analysis of this fragment. In order to analyze the structure of hTAF2 core in more detail, a series of limited proteolysis experiments was done on the 116 kDa fragment of the hTAF2 TEV construct A. This fragment was the biggest metastable fragment in the proteolytic digests of the full-length hTAF2 protein, and since it is smaller than the full-length hTAF2 (137 kDa), the analysis of its proteolytic fragments is somewhat easier. The final goal of this analysis was to determine stable, well structured pieces, that could be produced separately and analyzed structurally and functionally. More detailed determination of the disordered regions in this construct would allow design of the better hTAF2 crystallization constructs.

As a first step, a comparison was made between fragments resulting from 3 different protease digests: trypsin, elastase and subtilisin. Since these proteases have different specificities (Figure 4.6), fragments that appear stable in different digests could be assigned as separate domains with more certainty. All digests were performed with the purified hTAF2 construct A

(116 kDa fragment) concentrated to 2 mg/ml. Samples were prepared for the SDS-PAGE analysis as described in the section 2.6.6. Trypsin was added to the mass ratio of 1:1000, like for the full-length hTAF2 analysis (section 4.2.3), and samples were taken after 1, 5, 15, 30, 60, 120 and 240 min. Elastase and subtilisin showed better results (fragments over wider molecular weight range and therefore more informative) if they were used in different concentrations instead of performing a time course with one concentration. Elastase was used in protease:protein ratios from 1:500 to 1:5, and subtilisin, which was more active than elastase, from 1:20000 to 1:1000. Comparison of tryptic, elastase and subtilisin fragments (Figure 4.24 and Figure 4.25) revealed two very stable fragments in common to all proteases, with approximate molecular weight of 60 kDa and 20 kDa. All proteases also revealed a 25 kDa fragment that was somewhat less stable than the first two.

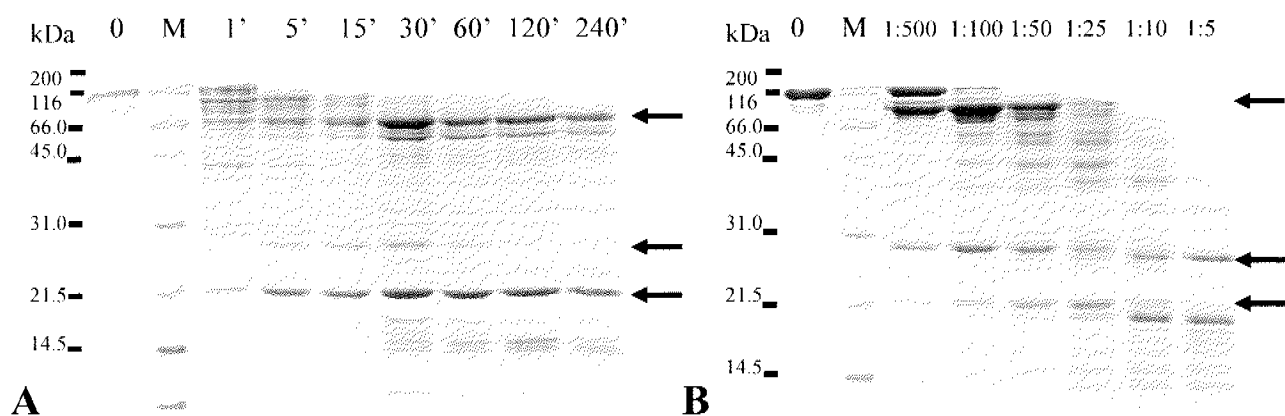


Figure 4.24. Limited proteolysis of hTAF2 core domain: trypsin and elastase fragment comparison. A) Time course of the 1: 1000 (mass ratio) digest with trypsin. B) Elastase digest with different protease concentrations (30 min). Mass ratios are indicated. The arrows point to 3 common stable fragments, about 60, 25 and 20 kDa in size. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

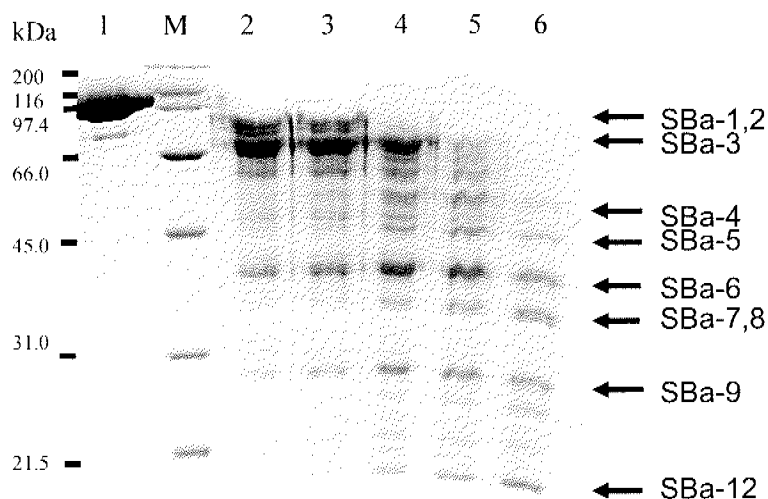


Figure 4.25. Limited proteolysis of hTAF2 core domain with subtilisin.

Lane 1- sample before protease addition, lane 2- 1: 20 000, lane 3- 1: 10 000, lane 4- 1: 5 000, lane 5- 1: 2 500, lane 6- 1: 1 250 (mass ratio of protease: protein). All samples were digested for 5 min. Lane M contains the broad range SDS-PAGE protein marker (BioRad). Fragments whose analysis was attempted by N-terminal sequencing are marked with arrows. SBa-3, SBa-9 and SBa-12 correspond to three stable fragments common to both the tryptic and elastase digests.

In order to identify the fragments, MALDI and N-terminal sequencing analysis was performed with the tryptic fragments, since their SDS-PAGE separation appeared to be the best (Figure 4.26, Table 4.3). Most of the fragments could be identified with high confidence, their calculated molecular weight from the N-terminal sequencing data and SDS-PAGE estimated size corresponded well to the MALDI data. High molecular weight fragments TRa-1 and TRa-1a were more difficult to identify confidently, as their MALDI signal to noise ratio was low. The exact C-terminus of these fragments was predicted by taking into account the other fragments and identified proteolytic sites. Another technical difficulty comes from the fact that these bands are a mixture of 2 or more fragments (one having an intact N-terminus and the other starting from position 12, where the sequence was determined).

Fragment	Size (kDa)	N-terminal sequence	Start (aa)	End (aa)	Sequencing quality
TRa-1	92.2 ?	rKKGD	12	815?	acceptable
TRa-1a	83.0 ?	rKKGD	12	735?	good
TRa-2	66.9	rKKGD	12	595 and 596	good
TRa-3	56.7	rNLNYF	100	596	acceptable
TRa-3a	46.6	rKKGD	12	420	bad (heterogeneous)
TRa-4	25.3	rNKKKK	597	815	good
TRa-5	22 ?	rTLDNL	816	999?	good
TRa-6	20.9	rTLDNL	816	981?	good
TRa-7	16	rNKKKK	597	735	good
TRa-8	14.1	rNKKKK	597	718	good
TRa-9	6.1	kTNNFM	736	786	good

Table 4.3. Results of MALDI analysis and N-terminal sequencing of TAF2 core domain tryptic fragments.

The N-terminal sequence, as obtained by Edman degradation, is shown in capital letters. The final amino acid (“End”) is estimated from the fragment size determined by MALDI, position of the potential tryptic cutting sites and the relation to other fragments. Question marks mark the fragments where the MALDI mass didn’t correspond exactly to the expected mass calculated from the end tryptic site position and for which the signal to noise ratio was low.

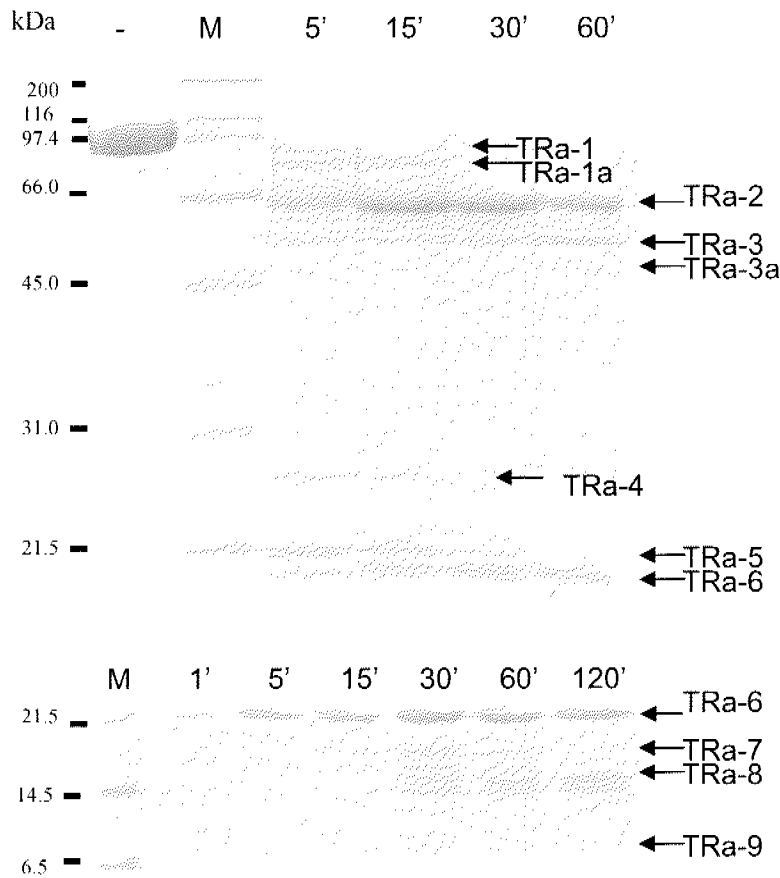


Figure 4.26. Time course of the hTAF2 core domain tryptic digest (1:1000 mass ratio). Higher and lower molecular weight range are analyzed by 12% (top) and 15% SDS-PAGE (bottom). The arrows point to the tryptic fragments analyzed by MALDI and N-terminal sequencing. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Subtilisin fragment identification was attempted as well, but proved to be technically very difficult. The reason is that subtilisin is less specific than trypsin or elastase and produces a more diverse mixture of fragments, especially in the higher molecular weight range. Such fragments were impossible to separate well using SDS-PAGE. N-terminal sequencing was attempted with some of the bands, giving results of bad quality due to high heterogeneity (Table 4.4). Many lower molecular weight bands, which separated better, could not be sequenced because they didn't transfer to the PVDF membrane well.

Fragment	Size (kDa)	N-terminal sequence	Start (aa)	End (aa)	Sequencing quality
SBa-1	90	v VGFVE	43	end?	bad
SBa-2	80	a MVAVS	209	end?	bad
SBa-3	60	a MIIDE	332	end?	bad
SBa-4	50	could not be obtained	-	-	-
SBa-5	45	could not be obtained	-	-	-
SBa-6	40	m DADSP	618	end?	good

Table 4.4. Partial results of N-terminal sequencing with TAF2 core domain subtilisin fragments. The N-terminal sequence, as obtained by Edman degradation, is shown in capital letters. The fragment sizes are estimated by SDS-PAGE.

4.3.2.2 Analysis of endogenous proteolytic degradation fragments

Analysis of the endogenous degradation fragments of partially purified 2 months old hTAF2 construct A sample was done to compare the results obtained by limited proteolysis. There were 3 main degradation products (Figure 4.27), whose molecular weight was determined by MALDI to be 84, 29 and 20 kDa. N-terminal sequencing revealed that the bands were a mixture of several slightly different fragments whose signals could not be resolved in the MALDI analysis (Table 4.5). Taken together, these data clearly pointed to two regions (except the N- and C-termini) to be proteolytically the most sensitive: amino acid 737-739 and amino acid 817-818. They compared well to the detected tryptic sites (TRa-1a, TRa-7 and TRa-9 were cut at the position 735, while TRa-1, TRa-4, TRa-5 and TRa-6 were cut at the position 815). The fact that the fragment 1-817 doesn't appear (or not as strong as the others), implies that the region 737-739 is the most sensitive, producing fragments DEG-1 and DEG-2, followed by the further digest of the DEG-2 giving smaller fragment DEG-3.

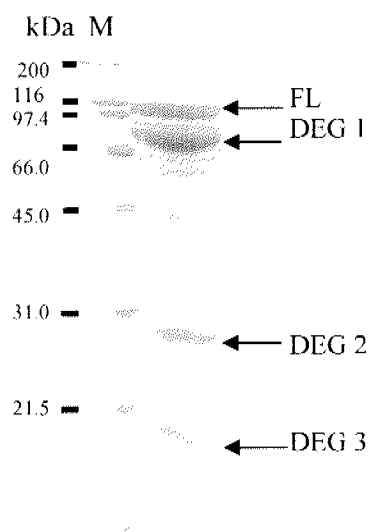


Figure 4.27. TAF2 core domain endogenous degradation fragments.

The sample is partially purified TAF2 construct A (116 kDa), kept for 2 months at 4° C. The arrows point to the positions of the full-length protein (FL) and 3 main degradation fragments. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Fragment	Size (kDa) MALDI	N-terminal sequence	Start (aa)	End (aa)	Sequencing quality
FL	113.4	not clear (multiple D, indicates N terminus)	~N-terminus (minus first 4 amino acids)	991	bad (heterogeneous sample)
DEG1	84.1 83.8	not clear (multiple D, indicates N terminus)	~N-terminus (minus first 4 amino acids)	737	bad (heterogeneous sample)
DEG2	29.3	3 sequences: NNFMS... NFMSQ... FMSQ...	737 738 739	990 991	good
DEG3	20.2	2 sequences: LDNLNP... DNLNP...	817 818	990 991	good

Table 4.5. Results of MALDI and N-terminal sequence analysis of TAF2 core domain endogenous degradation fragments. All possible fragment start and end positions are indicated.

4.3.2.3 Gel filtration experiments with proteolytic mixtures

Gel filtration runs were performed with the proteolytic fragment mixtures to get the information on their structural relationships. This analysis allows determining which fragments are co-folded and which can be purified separately. A special S200 column, designated only for proteolytic mixtures, was calibrated by a full-length hTAF2 (137 kDa) and hTAF2 construct A fragment (116 kDa). The full-length hTAF2 eluted with a peak at the fraction 26 (13 ml), while the construct A eluted at the fraction 28 (14 ml).

Preparative proteolytic digests were prepared by using 300 μ l of the 1.5 mg/ml purified hTAF2 construct A 116 kDa sample. For the tryptic digest, the sample was incubated with 1:1000 mass ratio of trypsin to protein, and for elastase, a 1:100 mass ratio was used. After 15 min of incubation, 3 mM PMSF was added to block the protease. A small precipitate was noticed after adding PMSF so the sample was centrifuged (12 000 rpm, 10 min, 4 deg, benchtop) before loading the column. The column was pre-equilibrated in the high salt buffer (25 mM HEPES-NaOH pH 7.6, 400 mM KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF), to make sure that weakly interacting protein fragments would be separated. After the run, which in the case of both trypsin and elastase resulted in only one peak, the peak fractions were 10x concentrated in the Amicon device and loaded on the SDS-PAGE for the analysis (Figures 4.28 and 4.29). Interestingly, all the proteolytic fragments, in both cases, were co-eluting with a peak at the fraction 28, just like the non-digested protein. The experiment was repeated with the tryptic digest using different sets of fragments resulting from digests stopped at different time points, but all the fragments were always co-eluting at the same peak position.

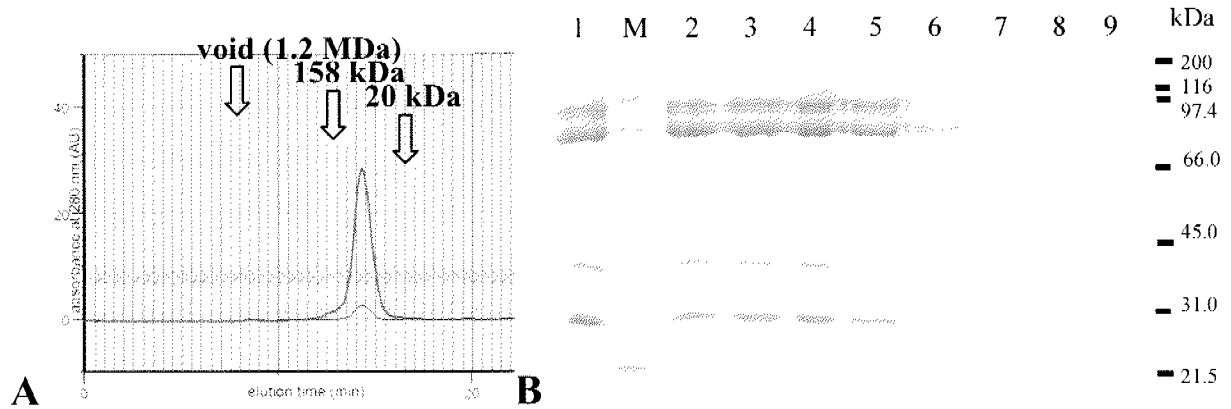


Figure 4.28. Gel filtration (S200 column) with tryptic degradation fragments of hTAF2 core domain. A) Chromatogram. Horizontal lines represent the 0.5 ml fractions (the peak corresponds to fraction 28). B) SDS-PAGE analysis of the fractions. Lane 1- column input, lanes 2 to 9- fractions 28 to 35. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

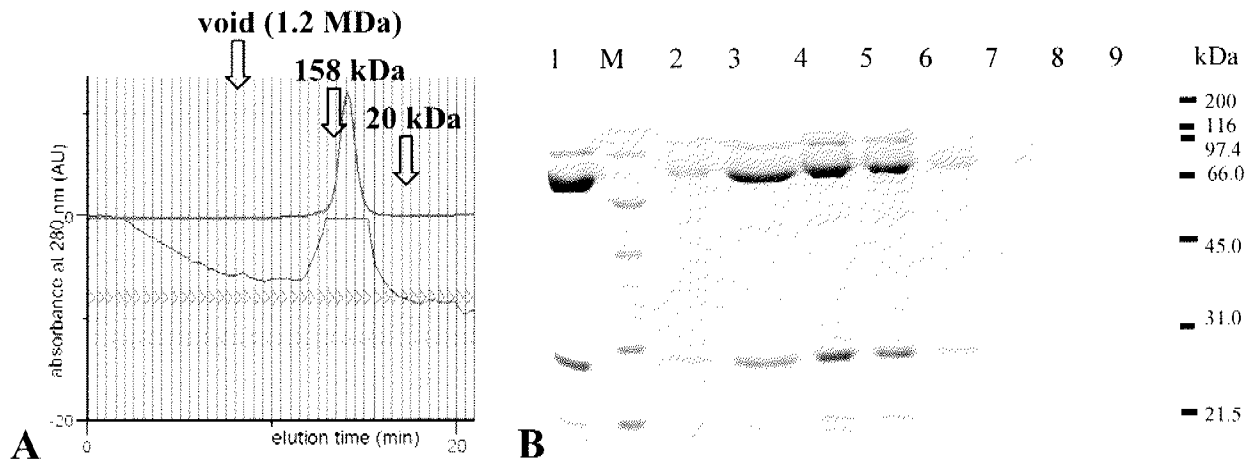


Figure 4.29. Gel filtration (S200 column) with elastase degradation fragments of hTAF2 core domain. A) Chromatogram. Horizontal lines represent the 0.5 ml fractions (the peak corresponds to fraction 28). B) SDS-PAGE analysis of the fractions. Lane 1- column input, lanes 2 to 9- fractions 27 to 34.

A series of experiments described in this section led to the conclusion about the hTAF2 core domain structure, which was determined to be contained roughly within its first 1000 amino acids. Figure 4.30 summarizes the results obtained with the core domain proteolytic fragment analysis, identifying the positions around amino acid 816, 597 and 736 as proteolytically the sensitive. Most importantly, gel filtration experiments with proteolytic fragments from trypsin and elastase digests showed that all detectable fragments are co-folded.

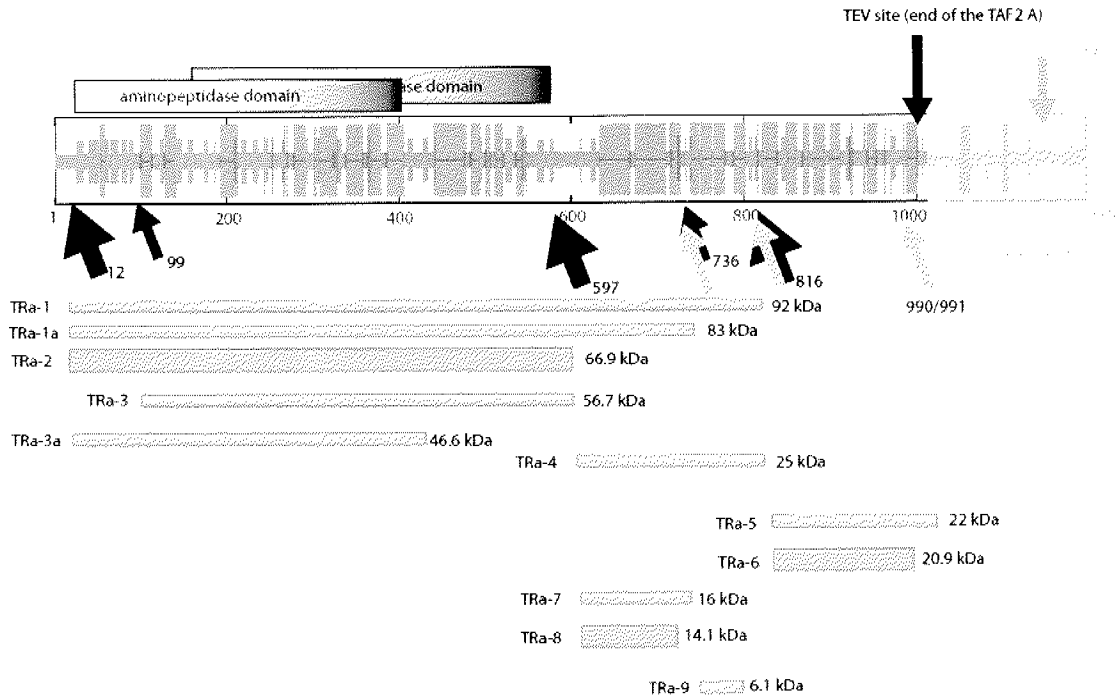


Figure 4.30. Summary of the limited proteolysis results and endogenous degradation of hTAF2 core domain.

hTAF2 full-length sequence (amino acid 1-1197) with the secondary structure prediction (α helix – blue, β strand – red, random coil – violet) is represented as a horizontal bar. The C-terminal tail domain, removed by a TEV digest, is shaded yellow. Tryptic hTAF2 fragments are shown below. Thickness of the fragment represents its apparent stability. Detected tryptic cutting sites are shown as a black arrow, with a number indicating the exact cutting position. The thickness of the arrow indicates the apparent cutting efficiency. Endogenous degradation positions are indicated by a violet arrow.

4.3.2.4 Ion-exchange chromatography with tryptic fragments

In an attempt to separate some of the fragments from the tryptic digest mixture, a series of ion-exchange and hydrophobic interaction chromatography experiments was performed. First, solubility and binding tests with the non-digested TAF2 construct A fragment were performed in order to determine the appropriate buffer conditions. For the solubility test, 0.3 mg/ml protein aliquots were dialyzed in the 25 mM HEPES-NaOH pH 7.6 buffers containing 10 mM, 25 mM, 1.5 M or 2 M KCl. Samples were centrifuged (12 000 rpm, 10 min, 4 deg, benchtop) and analyzed by SDS-PAGE (Figure 4.31). There was no visible pellet either in the tube or on the gel, so the conclusion was made that TAF2 construct A is stable in a range between 10 mM and

2 M KCl. Higher ionic strength buffers were tested as well, but the protein precipitated in 2 M ammonium sulphate and above.

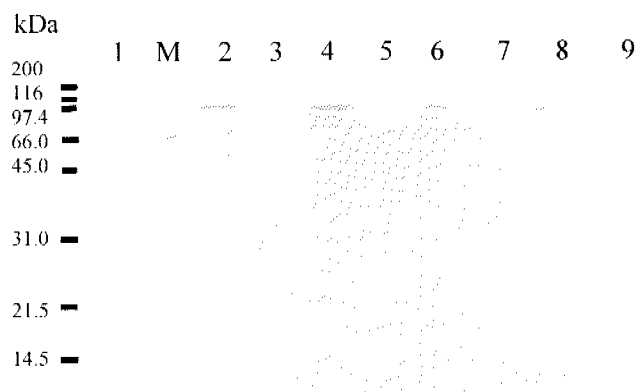


Figure 4.31. TAF2 construct A solubility test in a low and high salt range.

Lane 1- sample before dialysis, lanes 2, 4, 6 and 8 – supernatant after dialysis into 10 mM, 25 mM, 1.5 M and 2 M KCl-containing buffer, respectively. Lanes 3, 5, 7 and 9 – corresponding pellets. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

The next step was to determine the pH and salt conditions that will allow the binding of the sample to the Resource S (cation exchange), Resource Q (anion exchange) and Resource Phe (hydrophobic interaction) columns. Different buffers were extensively tested, but only the binding to the Resource S column could be established. Since the predicted pI of the construct A was 7.3, the pH of the buffer had to be at least one unit lower to allow the binding. A sample of the non-degraded construct A was prepared by diluting an aliquot in the storage buffer 20x with the Buffer A (50 mM MES pH 6.0, 1 mM PMSF, 1.5 mM DTT) and then dialyzing it for 2 h in the same buffer. The sample was loaded on the 1 ml Resource S column pre-equilibrated in the Buffer A and eluted with the linear gradient from 0-100% Buffer B (50 mM MES-HCl pH 6.0, 1M KCl, 1 mM PMSF, 1.5 mM DTT). The peak appeared at the KCl concentration of 200 mM. Finally, a trypsinized sample of construct A was prepared in the same way. Trypsin was added to the mass ratio of 1:1000 and the sample was incubated for 15 min. Digest was stopped with 3 mM PMSF and the sample was centrifuged before loading on the Resource S column. The chromatogram (Figure 4.32 A) showed a double peak at the position where only a single peak with the non-digested sample appeared in the previous experiment. The salt gradient had to be flattened at the double peak position to allow the peak separation. Peak fractions were

concentrated using the Amicon device and analyzed by SDS-PAGE (Figure 4.32 B). Depending on the tryptic digest length, the intensity of the two peaks varied. The first peak (A) increased with the longer digested sample without TRa-1 and TRa-1a fragments, while the second peak (B) contained only the TRa-1, TRa-1a, TRa-2 and TRa-6 fragments. This result indicated that there is a peptide, contained within TRa-1 and TRa-1a fragments but missing from peak A fragment mixture that is responsible for different ionic properties of these two peaks. The analysis of the fragments pointed to the proteolytically sensitive region between the end of TRa-8 and beginning of TRa-6 fragments.

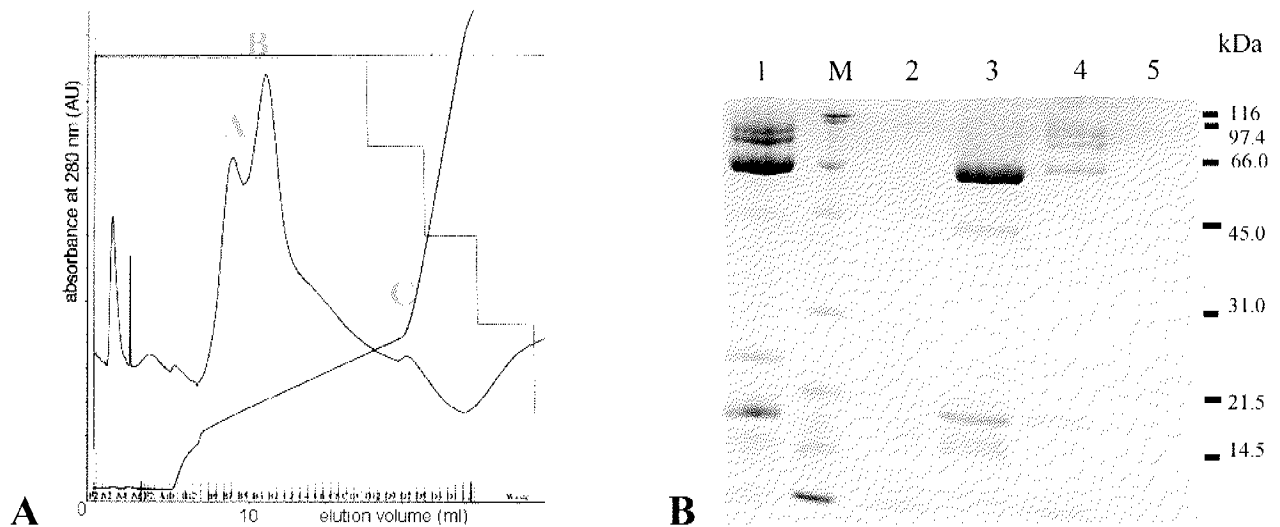


Figure 4.32. Cation-exchange chromatography (Resource S column) with tryptic fragments of the hTAF2 core domain.

A) Chromatogram. Three main peaks are marked A, B and C. B) SDS-PAGE analysis of the concentrated peak fractions. Lane 1- input sample, 2- flowthrough fraction, 3- peak A, 4- peak B, 5- peak C. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

4.4 Production and characterization of hTAF2 tail domain

hTAF2 tail domain was defined in limited proteolysis experiments of the full-length protein (section 4.2) as a proteolytically sensitive and therefore disordered part of TAF2. It consists of amino acids 1000-1197 (starting with the last detected tryptic position and ending at the full-length TAF2 C-terminus) and contains a low complexity region rich in histidines and lysines, which make a cluster of positive charge near the end of the sequence (Figure 4.37). This domain could be produced as a part of the TAF2 TEV construct A (expressed in Sf21 insect cells) and released from the core domain by TEV protease cleavage (section 4.3.1.2). Even though it is possible to purify it by this strategy, the yield was quite low, so it was decided to express this domain separately in *E. coli* and use this material for further structural and functional studies.

4.4.1 Cloning, bacterial expression and purification of hTAF2 tail domain

hTAF2 tail domain was cloned into the *E. coli* pET28a expression vector by using *NcoI* and *XhoI* restriction enzymes. PCR was performed over the desired region of hTAF2 gene using the pDiFBhTAF2 plasmid as a template and TAF2tailN and TAF2tailC oligonucleotides (containing the *NcoI* and *XhoI* restriction sites, respectively) as primers. The resulting clones were analyzed by *NcoI* / *XhoI* double digest and confirmed by sequencing.

A positive clone containing the hTAF2 tail gene was used for transformation of *E. coli* Rosetta expression strain. The cells were grown in a small scale liquid culture to OD₆₀₀ of 0.5 and induced with 0.42 mM IPTG. The induced cultures were incubated for 3 h at 37° C. Samples of induced and uninduced control culture were taken for the SDS-PAGE analysis (Figure 4.33 A, lanes 1 and 2), which clearly showed the expressed band. Interestingly, the band migrated more slowly on the SDS-PAGE than expected from its molecular weight (22 kDa), just like the same fragment released from the TAF2 TEV construct A expressed in insect cells (Figure 4.15). The reason is the high pI of this fragment (predicted to be 9.7).

To check the solubility of the expressed tail domain, the cells were resuspended in the Bacterial lysis buffer containing the 50 mM imidazole. After 2x20'' sonication, the cells were centrifuged in the preparative ultracentrifuge (40 000 rpm, 60 min, 4° C, Ti-70). The supernatant and the pellet were analyzed by SDS-PAGE (Figure 4.33 A, lanes 3 and 4). To test if the

expressed protein can be purified using the internal histidine stretch (used also for full-length TAF2 purification), the supernatant was incubated with TALON™ resin (batch purification) pre-equilibrated in 50 mM imidazole-containing Bacterial lysis buffer. The resin was washed with the same buffer containing 1 M KCl and the protein was eluted using 200 mM imidazole-containing buffer (Figure 4.33, lanes 5, 6, 7). The expressed protein could be purified from the soluble fraction, but it had degradation problems. Even if all available protease inhibitors are added to the buffers and purification is done fast at 4° C, at least one degradation band (DEG-1) was always present in the TALON™ elution sample. Older concentrated samples revealed even more degradation products (Figure 4.33 B) that are probably appearing already in the expression stage. This result is not surprising, since the tail domain is proteolytically the most sensitive part of TAF2 and predicted to be poor in secondary structure content.

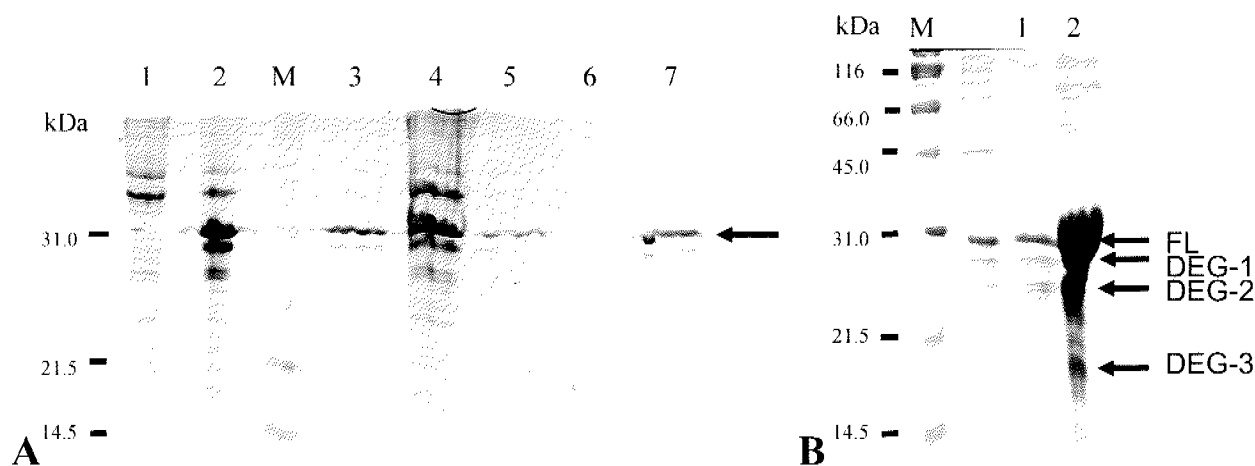


Figure 4.33. Expression and TALON™ batch purification of hTAF2 tail domain.

A) Lane 1- non-induced cells, 2- cells after 3 h IPTG induction, 3- TALON™ resin input (protein extract supernatant), 4- concentrated pellet from protein extract, 5- unbound sample, 6- wash fraction, 7- elution fraction. The arrow points to the main expressed band position. B) Lane 1- old sample with degradation products, 2- concentrated old sample. Arrows point to the full-length protein and 3 main degradation fragments. Fragment DEG-2 is a double band. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

An attempt was made to purify the full-length tail domain from the degradation products using an imidazole gradient in TALON™ chromatography or a salt gradient in cation-exchange chromatography. TALON™ chromatography was performed with a 2 ml low pressure column pre-equilibrated in TALON™ buffer A (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 10 mM imidazole, 1 mM PMSF, 10 mM 2-mercaptoethanol). A linear gradient was formed by using a 0-

100% concentration of TALON™ buffer B (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 200 mM imidazole, 1 mM PMSF, 10 mM 2-mercaptoethanol). The sample was prepared like for the TALON™ batch purification, with the difference that Buffer A was used as lysis buffer. The results (Figure 4.34 A) showed that the DEG-2 fragment had slightly different elution properties than the full-length domain and could be only partially removed using this procedure. Since it eluted with a lower imidazole concentration than the rest of the mixture, this fragment might have lost some of its histidine residues while still retaining enough to be able to bind to the resin. Cation-exchange chromatography was performed using the 1 ml Resource S column pre-equilibrated in Buffer A (25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 1 mM PMSF, 10 mM 2-mercaptoethanol), which was used also as a lysis buffer. After the sample was loaded, the elution was performed using the linear gradient of 0-100% Buffer B (25 mM HEPES-NaOH pH 7.6, 1 M KCl, 1 mM PMSF, 10 mM 2-mercaptoethanol). The results (Figure 4.34 B) showed that the DEG-2 fragment can be efficiently purified from the rest of the mixture since it elutes with the lower salt concentration, implying that this fragment has lost also some of the positively charged residues that contribute to the binding. Unfortunately, none of these procedures were useful in obtaining the pure full-length TAF2 tail domain without the most prominent DEG-1 degradation fragment.

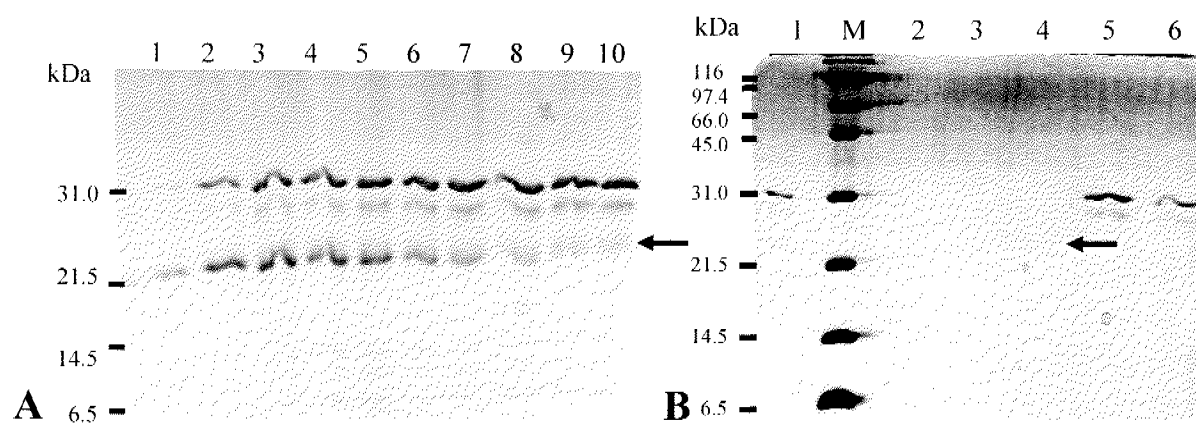


Figure 4.34. Attempts to separate the TAF2 tail domain from degradation fragments. A) TALON™ chromatography. 10 – 200 mM imidazole gradient was used. Lanes 1-10 show the peak fractions. The arrow points to the DEG-2 lower band that elutes slightly before the full-length and DEG-1 fragment. The marker bands are indicated on the side. B) Resource S chromatography. 0.1 – 1 M KCl gradient was used. Lane 1- input sample. Lanes 2-6 show the peak fractions. The arrow points to the DEG-2 band which can be efficiently separated from the rest since it elutes at lower salt concentration. Lane M contains the broad range SDS-PAGE protein marker (BioRad).

Gel filtration analysis was performed with the batch TALON™ purified TAF2 tail domain to check its oligomeric state. For this purpose, the S200 column was pre-equilibrated in the buffer containing 25 mM HEPES-NaOH pH 7.6, 200 mM KCl, 1 mM PMSF and 10 mM 2-mercaptoethanol. The resulting chromatogram (Figure 4.35) had 3 peaks, one corresponding to the void volume fraction. This peak contained no detectable protein, so the conclusion was made that it probably comes from the nucleic acids in the sample (the 1 M KCl TALON™ resin wash was not performed for this experiment, so some DNA could have stayed on the resin and eluted along with the protein). A small peak A contained the fragment that would correspond to the double molecular weight of the tail domain, indicating it might come from its disulfide-linked dimer (the tail domain contains one Cys residue). The large peak B contained most of the sample. From this gel filtration experiment it could be concluded that the TAF2 tail domain exists in the monomeric state.

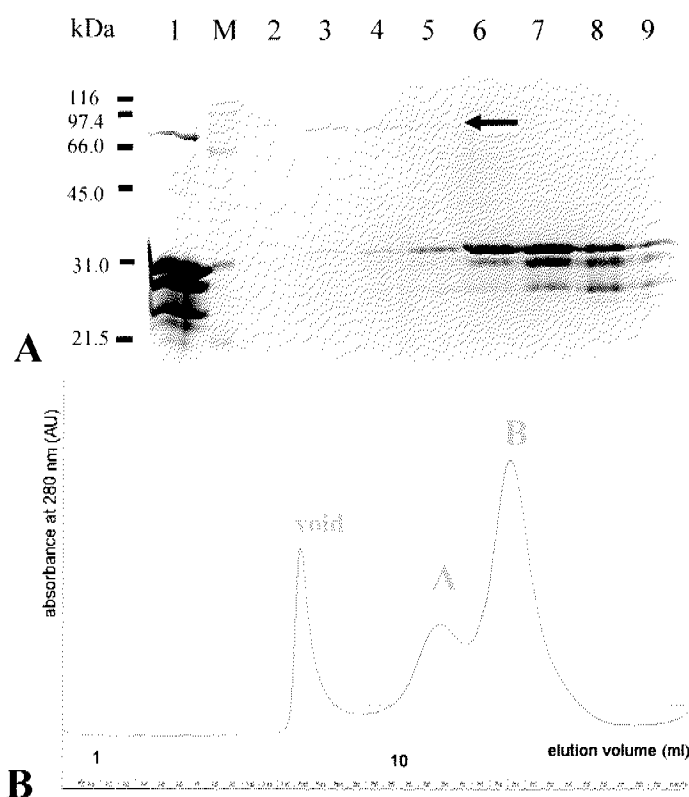


Figure 4.35. Gel filtration with the TAF2 tail domain (S200 column).

A) SDS-PAGE analysis of the fractions. Lane 1- input sample, 2- void volume peak, 3 to 9- fractions B4, B3, B3, B2, B1, C1, C2, C3 (peaks A and B). The arrow points to the suspected disulfide dimer band.
B) Chromatogram showing 3 peaks. The sample is contained within peak B.

4.4.2 Analysis of endogenous degradation fragments

To identify the endogenous TAF2 tail domain degradation fragments, MALDI and N-terminal sequencing analysis was performed. Figure 4.36 shows MALDI results that include all prominent degradation fragments, as well as the full-length protein. The results of N-terminal sequencing and fragment masses obtained by MALDI analysis are listed in the Table 4.6. DEG-1 fragment, with the peak molecular weight of 19.96 kDa, probably consists of a fragment mixture with a heterogeneous C-terminus, judging from the peak shape. DEG-2 fragment, a double fragment by the SDS-PAGE analysis, consisted of two fragments with a C-terminal 4 amino acid difference and corresponding masses of 17.1 and 17.6 kDa (17.1 kDa fragment is more prominent on the gel). Different chromatographic properties of the 17.1 kDa DEG-2 fragment were also taken into account during fragment identification. They can be explained by the loss of some lysine residues and a HK repeat sequence that are contributing to the strength of Resource S and TALON™ column binding, respectively. 17.6 kDa DEG-2 fragment retained this sequence and therefore behaved chromatographically the same as the full-length fragment. DEG-3 fragment could not be identified with high confidence since its N-terminal sequence showed high heterogeneity.

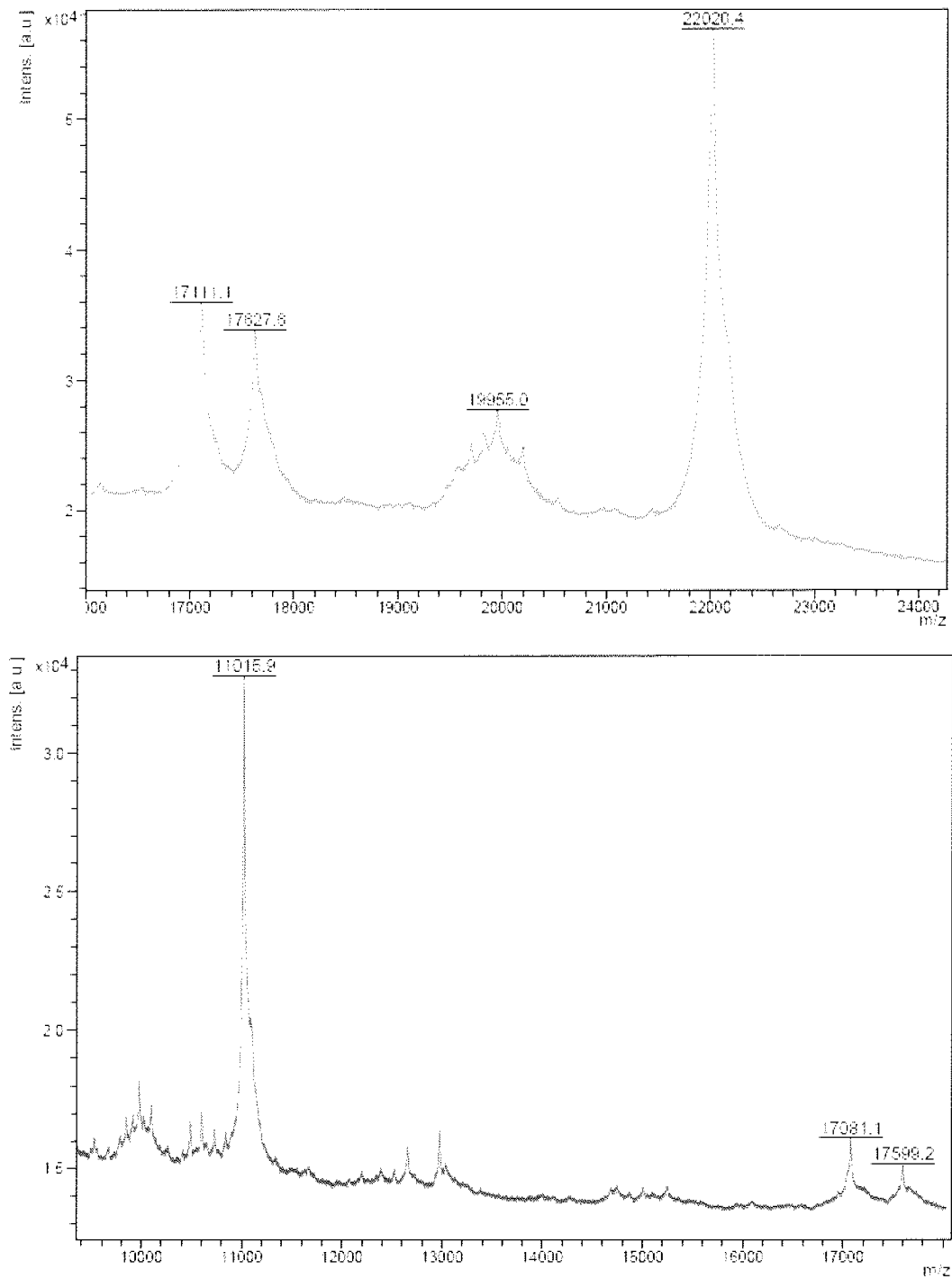


Figure 4.36. MALDI analysis of hTAF2 tail natural degradation fragments. 5 main fragments are detected by MALDI; 22 kDa corresponding to the full-length tail domain and 4 degradation products.

Fragment	Size (kDa) MALDI	N-terminal sequence	Start (aa)	End (aa)	Sequencing quality
FL	22.0	GKAVL	2	199	good
DEG1	19.9	GKAVL	2	179 (heterogeneous)	good
DEG2 (double band)	a) 17.6 b) 17.1	GKAVL	2	a) 161 b) 157	good
DEG3	11.0	MDTVH (heterogeneous)	46	148	bad

Table 4.6. Results of MALDI and N-terminal sequence analysis of TAF2 tail domain natural degradation fragments.

Figure 4.37, showing the multiple sequence alignment and the secondary structure prediction of the TAF2 tail domain, summarizes the proteolytically most sensitive regions.

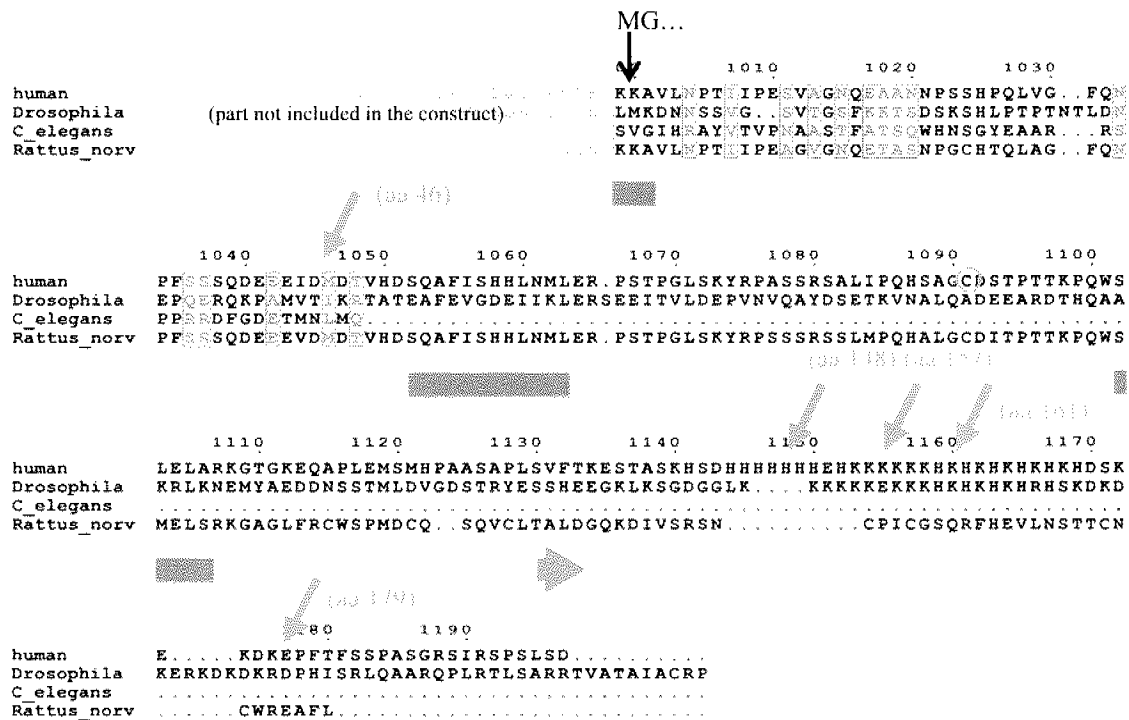


Figure 4.37. Summary of hTAF2 tail natural degradation fragment analysis.

The alignment of human, *Drosophila*, *C. elegans* and *R. norvegicus* TAF2 tail sequences is shown, with the predicted secondary structure elements (two helices). The detected proteolytic sites are indicated by red arrows. The human sequence used in this work starts with MGKAVL... . The only cysteine residue in the human sequence is circled.

4.4.3 Structural analysis by CD spectroscopy

Circular dichroism (CD) spectroscopy analysis was performed as described in the section 2.7.2 in order to measure the secondary structure content of the hTAF2 tail domain (residues 1000-1197). This domain was predicted to be largely disordered by different theoretical methods (section 4.2.1), and the experimental results obtained by limited proteolysis of hTAF2 indicated the same. hTAF2 tail domain sample was prepared for the CD analysis by TALON™ batch purification (described in the section 4.4.1) followed by dialysis into the CD buffer (10 mM Tris-Cl pH 8.0, 100 mM KF, 2 mM MgCl₂ and 2 mM 2-mercaptoethanol). The buffer was exchanged 4 times in the course of 6 h to make sure most of the chloride ions are removed (they produce a strong CD signal, especially at lower wavelengths). A CD spectrum was recorded with the 0.27 mg/ml TAF2 tail domain sample. Figure 4.38 shows the final results of the measurement (CD signal with buffer signal subtraction converted to the units of molar ellipticity). TAF2 tail domain clearly shows a minimum below 200 nm, typical for disordered structures. There is also a weak negative signal between 220 and 230 nm that indicates the presence of some alpha helical structures. The other typical alpha helix minimum (208 nm) might be obscured by the strong disordered structure signal. Therefore, the CD spectrum of the TAF2 tail domain confirmed the previous theoretical and experimental results.

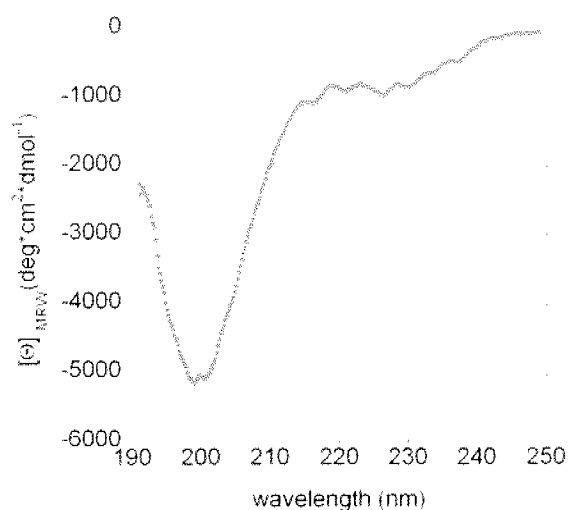


Figure 4.38. hTAF2 tail secondary structure analysis by circular dichroism.

The signal minimum characteristic for the denatured structures is visible at 198 nm. Existence of some helical structures can not be excluded since there could be a minimum at 208 and 222 nm as well.

4.5 Discussion

The series of experiments described in this chapter delineated a likely hTAF2 structural core domain constituted from approximately the first 1000 amino acids of the full-length protein. Results from proteolysis analysis of hTAF2 clearly showed that the last 197 amino acids are not structurally compact since they were very sensitive to different proteases. This region of the protein was named “TAF2 tail”, describing its floppy nature. The question whether the rest of the protein consists of several structurally separate domains or represents one compact core domain was answered by gel filtration experiments. Samples of proteolytically digested hTAF2 amino acid 1-1000 fragment were all co-eluting when loaded on the S200 column, even in high salt (400 mM KCl) conditions. This very important result showed that first 1000 amino acids of hTAF2 form a compact structure, in which proteolytically stable pieces are co-folded. This part of hTAF2 was therefore named “TAF2 core”. The analysis of six TEV constructs (section 4.3.1), initially designed for production of potentially crystallizable structural domains of hTAF2, confirmed these results. The construct A allowed the purification of two hTAF2 fragments resulting from the TEV digest: a 116 kDa fragment (the largest fragment stable in the trypsin, elastase and subtilisin digests) and the 22 kDa tail domain (last 197 amino acids). The construct B had the TEV site at the second proteolytically most sensitive position (amino acid 816). Even though TEV could successfully cut at that position, the two fragments could not be separated by TALON™ chromatography or other treatments (including high salt and mildly denaturing conditions), indicating strong co-folding of the amino acid 816-1000 region with the amino acid 1-815 part of hTAF2. Therefore, amino acid 816-1000 could not constitute a separate structural domain of hTAF2. This conclusion was confirmed by the analysis of the construct AB, where the “tail” domain (amino acid 1000-1197) could be purified away from the other two pieces (amino acid 1-816 and amino acid 816-1000). Constructs with the TEV site at position D (construct D, AD and D2) did not provide additional information on the domain structure since the TEV protease could not cut at positions D (amino acid 581) or D2 (amino acid 596). This result indicated that the inserted TEV site was inaccessible. Interestingly, this region (amino acid 580–620) is highly conserved, even though it doesn’t appear to have any secondary structure elements (Figure 4.1). The fact that it is only mildly proteolytically sensitive in comparison to the tail region, considering its length and large number of potential tryptic sites, indicates that most of

this long loop is somehow buried within the rest of the structure. That would explain the fact that the TEV protease recognition site inserted in this region is inaccessible to TEV. Interestingly, D, AD and D2 TEV constructs did not appear to have a strongly disrupted structure due to the TEV site insertion in this region, since they behaved exactly the same chromatographically as the wild-type hTAF2 protein.

Detailed limited proteolysis analysis was performed on the TAF2 core to obtain information on disordered parts of the structure, since the TAF2 core, just like the full-length TAF2, failed to produce crystals (Appendix III). Useful information was obtained from these experiments that can guide future crystallization efforts. Generally, the results on the full-length hTAF2 and the core domain overlapped well, pointing to several disordered or exposed regions: the most sensitive part of the core domain was the loop with the tryptic site at the position 816, consisted of 17 non-conserved residues that could be potentially removed from the crystallization construct. The N-terminus, containing the FLAG-tag with an enterokinase cleavage site, was sensitive as well. The other sensitive region was around residue 597, which is within a very conserved 40-residue loop. It is not clear if this loop represents a floppy region that should be removed from the crystallization construct: results from the TEV construct D and D2 analysis and its conserved character indicate that this region might be structurally important. Residue 736 was also identified as a tryptic cutting site, as well as the most sensitive position judging from the endogenous degradation fragments. Taken together, tryptic fragments TRa-2, TRa-6 and TRa-8 were the most stable and appeared to be co-folded based on the gel filtration experiments. Cation exchange chromatography with tryptic digest mixtures provided another clue that might be important for crystallization (Figure 4.32): mixtures missing a peptide somewhere from the region 736-816 could be purified away from the mixtures containing this peptide (since it is a part of the TRa-1 and TRa-1a fragments). This peptide does not appear to belong to the structural core and, when identified correctly, might hold a key to TAF2 crystallization.

TAF2 tail domain was identified by limited proteolysis of the full-length protein as the C-terminal disordered region of roughly 200 amino acids. The disordered nature of this domain was confirmed by secondary structure content measurement using the circular dichroism spectroscopy, showing a minimum below 200 nm typical for denatured proteins. According to the secondary structure prediction programs, this region might have two alpha helices: the CD spectrum doesn't exclude this possibility since two minima characteristic for helices might be

partially occluded by the large disordered minimum signal. Degradation sensitivity of this domain (but not within the predicted helices) is another proof of its disordered structure, as well as its propensity to form disulfide-linked dimers. Nevertheless, apparently due to its high positive charge, it can be expressed solubly in *E. coli* and purified over a TALON™ resin. The presence of a seven histidine residues followed by a repeating HK sequence causes this domain to strongly bind to the TALON™ resin, eluting only with 150 mM imidazole. This fact was conveniently used for the purification of not only this domain, but also the full-length TAF2 and its TEV mutant versions.

Multiple sequence alignments of TAF2 sequences from different organisms (Figure 4.1) show a general drop in conservation after residue 1000, where the tail domain starts. Compared to the human TAF2 tail sequence, *C. elegans* sequence does not even have most of the tail domain (only first 50 amino acids which are not conserved). The yeast sequence is also rich in lysine residues, but otherwise there are no similarities. The rat sequence does not have multiple H and K at the end of the sequence, but the first 120 amino acids of this domain are almost identical to the human version. Zebrafish TAF2 sequence is almost identical in its entirety to human, with the exception of the first $\frac{3}{4}$ of the tail domain (it has the multiple H, K and HK sequences). Relation between human, rat and zebrafish sequences regarding the tail domain are therefore very interesting from the evolutionary point of view. It could be imagined that this domain is involved in fine tuning of the transcription activity through different modulators of TAF2 function. For example, some proteins could bind to TAF2 via the tail domain to regulate its function within TFIID. Data obtained during this work (section 5.4.2) show that the tail domain has DNA binding activity, possibly via its many lysine residues. The fact that *C. elegans* and rat TAF2 sequences are missing these lysine residues opens interesting questions about TAF2 function in these organisms and in the function of TAF2 DNA binding in general.

5

FUNCTIONAL CHARACTERIZATION OF hTAF2*5.1 Introduction*

TAF2 has no significant sequence similarity to any other proteins of known structure or function (section 4.2.1). Apart from being a subunit of the general transcription factor TFIID, there is little biochemical evidence for its function and interaction with specific TFIID subunits. Immunoprecipitation assays with the *Drosophila* TAF2 (dTAF2) indicated the interaction with dTAF1 and dTBP (Verrijzer et al, 1994), but the human homolog appeared to have no hTBP or hTAF1 binding activity in far-Western analysis (Kaufmann et al, 1998). Nevertheless, there was a report of recombinant hTAF1-dTAF2 complex assembled *in vitro* having a specific initiator DNA-binding activity (Chalkey and Verrijzer, 1999). A yeast TFIID protein-protein interaction map indicated weak interaction of TAF2 with TAF7 and TAF10 (Yatherajam et al, 2003), and there was a report of hTAF2 having TAF4 binding activity shown by far-Western analysis (Kaufmann et al, 1998). DNA binding activity was reported for *Drosophila* (Verrijzer et al, 1995) and human TAF2 (Kaufmann et al, 1998), but it was not clear if the binding is sequence specific. dTAF2 together with TBP were predominantly responsible for the TFIID DNA footprint. dTAF2 produced a distinct shift with the AdML promoter DNA alone or in complex with dTAF1 and dTBP (Verrijzer et al, 1994). Indirect evidence for hTAF2 core promoter binding activity was also provided by experiments which show that hTAF2 stabilizes TFIID binding to the core promoter in *in vitro* assays with partially purified components (Kaufmann et al, 1998). In addition to the report of hTAF1-dTAF2 having a specific initiator binding activity, experiments with purified recombinant hTAF2 indicated specific binding to the “GAG” sequence. In these random DNA selection experiments, hTAF2 could bind to the “GAG” sequence with five-fold greater affinity than to random sequences (Martin et al, 1999).

This chapter describes experiments performed in order to obtain direct biochemical evidence for hTAF2 interaction with hTAF1 and with hTAF8/10 subunits of TFIID, as well as to characterize the specificity of its DNA binding activity. The information obtained from the domain characterization of hTAF2 (Chapter 4) was used to localize these activities.

hTAF1 is a large protein of 1872 amino acids (213 kDa) which has multiple functions within the TFIID complex. Structurally, it stretches across two lobes of TFIID (Leurent et al.,

2004), having direct contact with TBP and possibly several other TAFs. TAF1 has not only a structural role but also several enzymatic activities that function in positioning and stabilizing TFIID on core promoters and alteration of chromatin structure to allow pre-initiation complex assembly. TAF1 contains a serine/threonine protein kinase domain which is involved in the phosphorylation of TFIIF α and TFIIA *in vitro*, ubiquitin-activating/conjugating activity and an acetyltransferase domain which can acetylate free and nucleosomal histones, TFIIE β and TFIIF. TAF1 binds activators which recruit TFIID to specific promoters, regulates binding of TBP to DNA, is involved in core-promoter, initiator-element recognition and binds acetylated lysine residues in core histones (Wassarman and Sauer, 2001). To structurally and functionally investigate the interaction of hTAF1 with hTAF2, a putative subcomplex of TFIID that was indicated to have a role in initiator element recognition (Verrijzer et al, 1995, Chalkey and Verrijzer, 1999), two strategies were taken in order to produce this complex recombinantly. The first strategy was the expression and purification of each component separately, followed by complex assembly *in vitro*. The second strategy was to co-express either full-length hTAF1 or its domains with the full-length hTAF2. Once the complex is produced, it would allow more detailed analysis of TAF1-TAF2 interaction domains.

TAF8/10 is a subcomplex of TFIID, consisting of TAF8 (35 kDa) and TAF10 (22 kDa) subunits that form a histone fold pair. TAF10 is known to be required for the expression of a different subset of genes essential for specific cell types and differentiation pathways (Metzger et al, 1999, Mohan et al, 2003, Indra et al, 2005). It has two other histone-fold domain-containing partners in addition to TAF8; TAF3 and SPT7L, and these partners regulate its nuclear import (Soutoglou et al, 2005). To investigate if this complex directly interacts with hTAF2, hTAF8/10 was recombinantly produced and purified. The interaction was analyzed using gel filtration and Biacore experiments.

5.2 Production of hTAF1 and attempts to analyze the hTAF1-hTAF2 interaction

5.2.1 Recombinant production and purification of full-length hTAF1

5.2.1.1 Analysis of hTAF1 primary structure by bioinformatics

Figure 5.1 shows the multiple sequence alignment with the secondary structure prediction of five TAF1 sequences: human, *D. melanogaster*, *C. elegans* (putative sequence), *A. thaliana* and hamster. As seen from the alignment, the most conserved, as well as the most structured part of TAF1 is the central region (amino acids 630-1638). Interestingly, *S. cerevisiae* has its TAF1 function divided in two proteins (Matangkasombut et al, 2000): yTAF1 (only 145 kDa in molecular mass, corresponding to the N-terminal half of higher eukaryotic TAF1) and Bdf1 (Bromodomain factor 1, corresponding to the C-terminal half, starting with the bromodomains).

hTAF1 sequence can be divided into several domains (Wassarman and Sauer, 2001):

1. N-terminal protein kinase (1-414)
2. Proline-rich (157-165)
3. HMG-box (1195-1273)
4. bromodomain 1 (1397-1467)
5. bromodomain 2 (1520-1590)
6. acidic (1627-1872)
7. C-terminal kinase (1425-1872)

It was predicted to be localized in the nucleus, with the potential nuclear localization signal in the region 1352-1358, and to be an unstable protein on its own. Secondary structure prediction shows many random coil regions, as does the disorder prediction using DisEMBL (described in the section 4.2.1 for the analysis of hTAF2) (Figure 5.2).

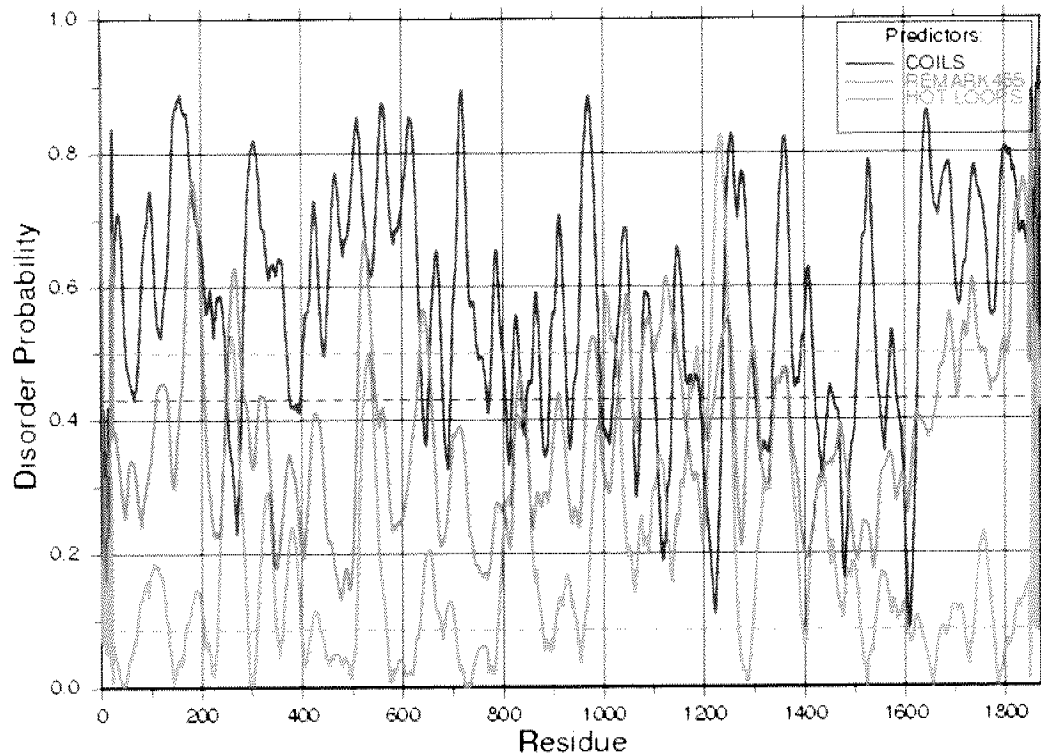


Figure 5.2. hTAF1 disorder prediction with DisEMBL.

Disorder probability, according to 3 different predictors, is shown in different colors. Dotted lines show a random expectation value.

5.2.1.2 Cloning of the hTAF1 expression construct

The gene for the full-length human TAF1 was obtained from Dr. I. Berger in the pUNI10TAF250 (pUNI10TAF1) cloning vector. To express this protein in Sf21 insect cells, it was necessary to subclone the gene in pFBDO expression vector, under the polyhedrin promoter. This vector encoded for a C-terminal His-tag for later protein detection and purification. To facilitate purification, additional tags were designed for the protein's N-terminus: calmodulin-binding peptide (CBP) and STREP-tag (streptavidin-binding peptide), followed by the enterokinase cleavage site (Figure 5.3).

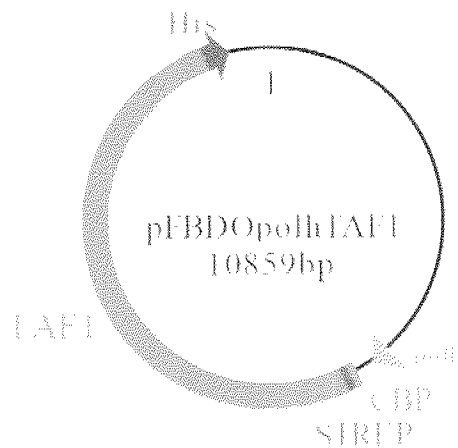


Figure 5.3. Map of the full-length hTAF1 expression vector pPBDOpolhTAF1. The full-length TAF1 gene in this vector contains a CBP-STREP tag fusion at its N-terminus followed by an enterokinase site and a His-tag at the C-terminus.

The hTAF1 gene is very large in size (5.4 kb) and special strategy had to be developed for its manipulation. 5'- and 3'-termini were therefore modified in two separate cloning steps. pUNI vector derivatives were very useful for the intermediate cloning steps since these vectors are small in size and have a low copy number in the appropriate host strain, reducing the amount of unwanted recombination products often seen when manipulating large genes like hTAF1.

In the first step, a *Rsr* II restriction site was inserted at the gene's 3'-end by a PCR reaction using the MM003 (with the *Rsr* II and *Bam*H I site) and MM004 (complementary to the gene region containing the *Bst*Z17 I site) primers ("e" and "d", Figure 5.4). The PCR product was ligated back into the original vector using the *Bst*Z17 I and *Bam*H I restriction sites. The resulting vector was named pUNI10h250Rsr.

In the second step, a large insertion coding for the N-terminal CBP-STREP tag had to be introduced. This was done using the gene-SOEing procedure (described in the section 2.2.20). A PCR was performed between the gene's 5'-end and a vector region containing the *Bss*H II restriction site. The resulting PCR product was ligated into the pUNI10h250Rsr vector using the *Nde* I (gene-SOEing primer encoded) and *Bss*H II (vector encoded) restriction sites. Primers MM006 and MM007 were used as reverse gene-SOEing primers ("a" and "b", Figure 5.4), together with MM005 ("c") as a forward primer. The resulting vector was named pUNI10BstCBPSTREP250Rsr.

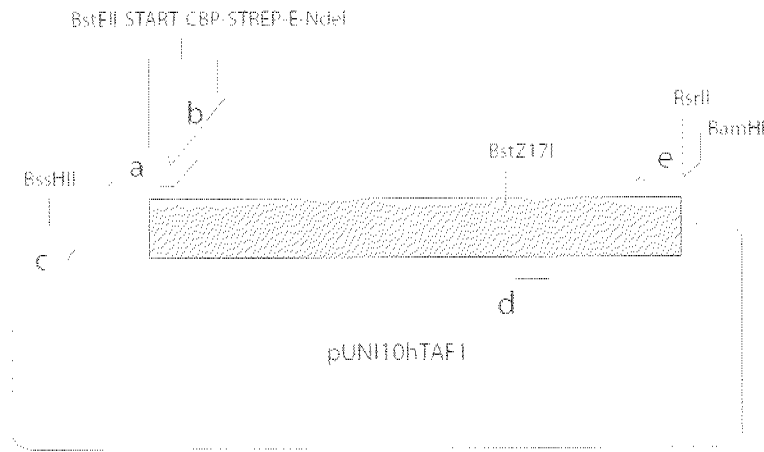


Figure 5.4. hTAF1 expression vector cloning strategy.

Finally, the *BstEII*–*RsrII* expression cassette was cut from the pUNI10BstCBPSTREP250Rsr vector and subcloned into the pFBDO expression vector to produce the pFBDOpolhTAF1 vector. The identity of the construct was confirmed by sequencing.

5.2.1.3 Expression of hTAF1 in Sf21 insect cells

The recombinant baculovirus containing the hTAF1 gene was produced according to the procedures described in the section 2.4. After the initial transfection, the virus was amplified twice before the expression could be detected using SDS-PAGE. Figure 5.5 shows the results of the expression (soluble fractions): the additional band, indicating hTAF1 expression, appeared approximately 48 h after the cells were completely infected, along with the late viral protein p10, and reached maximum level after about 72 h. The identity of the protein was confirmed by Western blot analysis using a His-tag antibody provided by Dr. I. Berger (Figure 5.5 B).

The large scale expression of the hTAF1 was done in 2 L shaker flasks and in the bioreactor, as described for the expression of hTAF2 (section 3.2.2). The expression level of the hTAF1 produced in the bioreactor was comparable to the levels obtained in the shaker flasks.

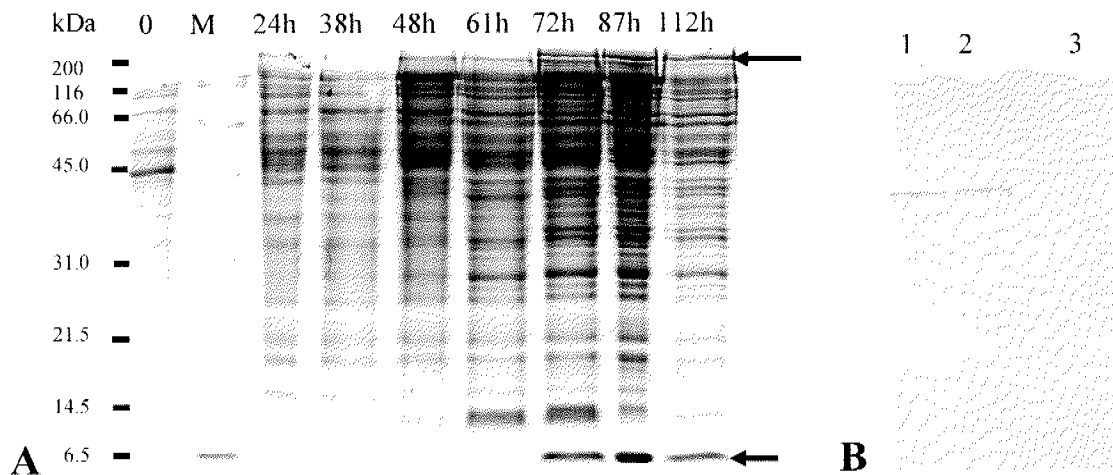


Figure 5.5. Expression of hTAF1 in Sf21 insect cells.

A) Time course of the viral infection. The big arrow indicates the position of hTAF1 band, appearing after 72 h. The small arrow points to the position of the late viral protein p10. Lane “0” contains the negative control sample and lane “M” contains the broad range SDS-PAGE protein marker (BioRad). B) Anti-His western blot analysis. Lanes 1 and 2- signal from two different TAF1-expressing viruses. Lane 3- non-infected cells.

5.2.1.4 Purification of hTAF1

In the course of this work, purification of recombinant hTAF1 was tried extensively with a variety of chromatographic resins and procedures. Only the most successful procedure will be summarized here. Generally, even though hTAF1 is expressed at high levels in Sf21 insect cells and present in the soluble fraction, the purification yield was low due to sensitivity to degradation and aggregation. hTAF1 appeared to have a strong DNA/chromatin binding affinity, a property that additionally complicated the purification procedure.

hTAF1 is predicted to be localized in the nucleus, according to the PSORT program (<http://psort.nibb.ac.jp>). This was confirmed experimentally by preparation of the nuclear and cytosolic fraction from Sf21 cells expressing hTAF1 (Figure 5.6). The cytosolic and nuclear protein extracts were prepared as described in section 2.5.1.2. This step was later used as the first purification step, allowing the separation of hTAF1 from the contaminating cytoplasmic proteins.

The pI of hTAF1, calculated using the ProtParam tool on the ExPASy server (<http://www.expasy.ch/tools/protparam.html>), was predicted to be 4.97, indicating a net negative charge for the protein at physiological pH values. Table 5.1 shows the primary structure analysis results and theoretical parameters of hTAF1 protein sequence.

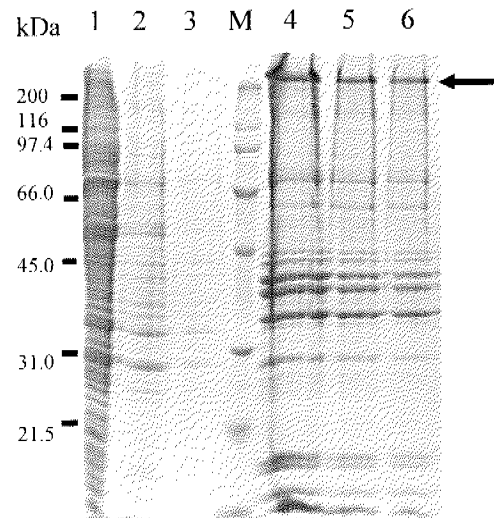


Figure 5.6. Cytoplasmic and nuclear fractions of Sf21 cells expressing hTAF1. Lanes 1 to 3- cytoplasmic fractions (from 3 consecutive resuspensions), 4- nuclear fraction (soluble and insoluble), 5- nuclear fraction after passing through the cell cracker, 6- soluble material from the nuclear fraction. The arrow indicates the position of hTAF1 band, present mainly in the nuclear fraction. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

Number of amino acids:	1872		
Molecular weight (Da):	212677.2		
Theoretical pI	4.97		
Amino acid composition:	Ala (A)	97	5.2%
	Arg (R)	106	5.7%
	Asn (N)	68	3.6%
	Asp (D)	145	7.7%
	Cys (C)	25	1.3%
	Gln (Q)	86	4.6%
	Glu (E)	193	10.3%
	Gly (G)	110	5.9%
	His (H)	35	1.9%
	Ile (I)	83	4.4%
	Leu (L)	152	8.1%
	Lys (K)	139	7.4%
	Met (M)	55	2.9%
	Phe (F)	61	3.3%
	Pro (P)	118	6.3%
	Ser (S)	143	7.6%
	Thr (T)	95	5.1%
	Trp (W)	18	1.0%
	Tyr (Y)	55	2.9%
	Val (V)	88	4.7%

Table 5.1. hTAF1 primary sequence analysis.

For the large scale purification of hTAF1, 5 ml of the Sf21 insect cell pellet expressing hTAF1 was resuspended two times in the Insect cell lysis buffer. The nuclei obtained after centrifugation (4500 rpm, 15 min, 4° C, tabletop) were resuspended in 20 ml of the Insect cell nuclei soak buffer, supplemented with 10 mM MgCl₂, 0.1 mM EDTA, 1.5 mM DTT and 1 mM PMSF. The nuclei were homogenized and passed twice through the cell cracker. The final volume of the nuclear extract was ~35 ml. The extract was centrifuged (18000 rpm, 60 min, 4° C, SS-34) and dialyzed overnight into Buffer A (25 mM HEPES pH 7.6, 200 mM KCl, 10 mM MgCl₂, 0.1 mM EDTA, 1.5 mM DTT, 1 mM PMSF and 10% glycerol), with one buffer exchange after 1 h. A minimum of 200 mM salt concentration was necessary to be present in all hTAF1 purification buffers since the protein precipitates completely at lower concentrations. All following purification steps were done at 4° C within 24 h. Purification at room temperature or over several days resulted in a significant increase in the amount of hTAF1 degradation products which were already clearly visible even if the procedure was done as fast as possible.

A 5 ml low-pressure Q-sepharose column (Pharmacia, 2.5 cm diameter) was prepared by a 5 column volumes wash with Buffer B (25 mM HEPES pH 7.6, 1 M KCl, 10 mM MgCl₂, 0.1 mM EDTA, 1.5 mM DTT, 1 mM PMSF and 10% glycerol), followed by a 5 column volumes wash with Buffer A. The Sf21 nuclear extract was loaded on the column using a peristaltic pump at the speed of 2 ml/min. After the sample was loaded, the column was washed with 20 column volumes of Buffer A. The following run was started: Buffer A for 2 ml, linear gradient from 0 to 30% Buffer B over 3 ml, linear gradient from 30 to 60% Buffer B over 20 ml, linear gradient from 60 to 100% Buffer B over 3 ml and 100% Buffer B for 2 ml. The resulting flow-through and elution fractions were analyzed by SDS-PAGE (Figure 5.7 A). Most of the hTAF1 was present in the flow-through fraction and did not bind to the column. A fraction of hTAF1 eluted in the center of the 30-60% Buffer B gradient, along with some impurities and most of the DNA from the extract. The run was repeated with the flow-through fraction, since it was purer than the elution fractions. Most of hTAF1 was now bound to the column (Figure 5.7 B).

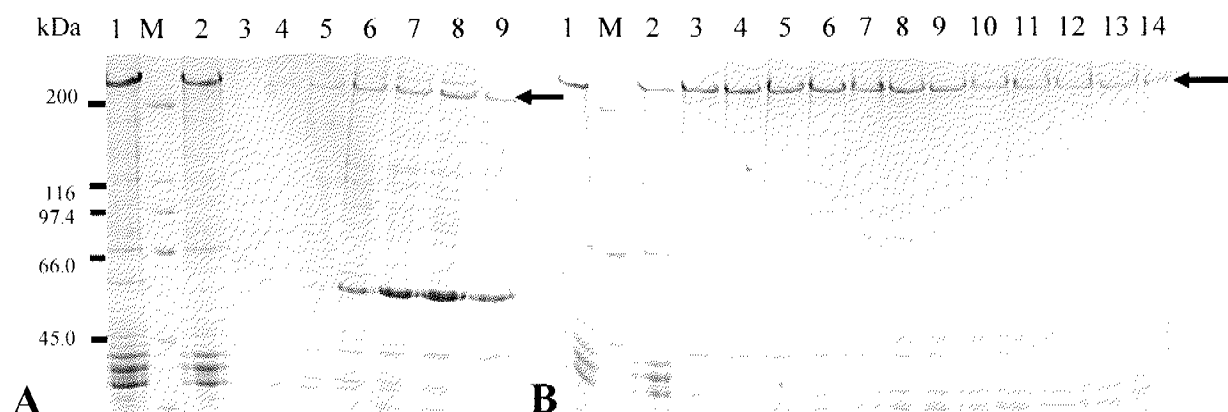


Figure 5.7. Q-sepharose column purification of hTAF1 from the nuclear fraction.

A) First Q-sepharose run. Lane 1- column input (nuclear fraction), 2- unbound fraction, 3-9 elution fractions. B) Second Q-sepharose run. Lane 1- column input (unbound fraction from the first run), 2- unbound fraction, 3-14 elution fractions. The arrows point to the position of the hTAF1 band. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

hTAF1 elution fractions from the second Q-sepharose column run were pooled and dialyzed against TALON™ Buffer A (25 mM HEPES pH 7.6, 400 mM KCl, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 1 mM PMSF and 10% glycerol). The dialysis was done in 4 h with one buffer exchange. The sample was centrifuged (12000 rpm, 10 min, 4° C, benchtop) and loaded on the 2 ml gravity-flow TALON™ column pre-equilibrated in the TALON™ Buffer A. After 10 ml wash using the same buffer, the samples were eluted with 4 ml of the TALON™ Buffer B (25 mM HEPES pH 7.6, 400 mM KCl, 10 mM MgCl₂, 200 mM imidazole, 10 mM 2-mercaptoethanol, 1 mM PMSF and 10% glycerol). 0.5 ml fractions were manually collected and analyzed on the SDS-PAGE. Figure 5.8 shows that all of the hTAF1 was bound to the resin and could be eluted. Most concentrated elution fractions still showed a significant amount of impurities, as well as the presence of hTAF1 N-terminal degradation products (bands immediately below the full-length protein in the high molecular weight range).

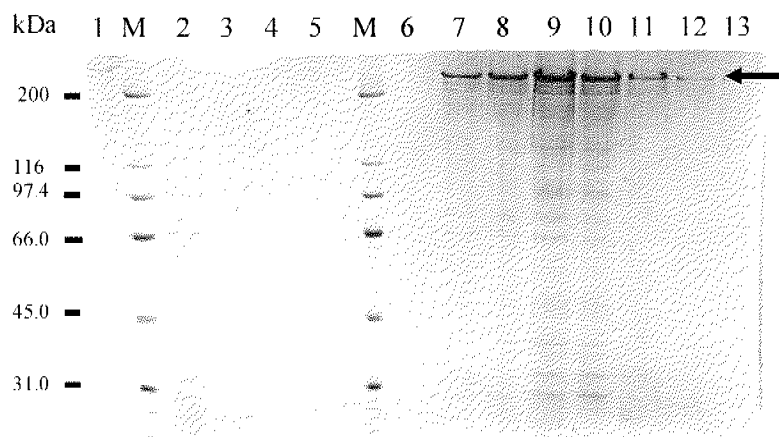


Figure 5.8. TALON™ purification with the Q-sepharose elution fractions of hTAF1. Lane 1- column input, 2- unbound fraction, 3 to 5- wash fractions, 6 to 12- elution fractions. The arrow points to the position of the hTAF1 band. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

Pooled TALON™ elution fractions were subsequently used for Calmodulin affinity resin purification to remove the contaminants as well as to pull out the non-degraded, full-length hTAF1. The sample was supplemented with 2 mM CaCl₂ and loaded on the 2 ml Calmodulin affinity resin gravity-flow column pre-equilibrated in the Binding buffer (25 mM HEPES pH 7.6, 400 mM KCl, 10 mM MgCl₂, 2 mM CaCl₂, 10 mM 2-mercaptoethanol, 1 mM PMSF and 10% glycerol). The wash was performed with 12 ml of the Binding buffer, followed by the elution using 4 ml of the Elution buffer (25 mM HEPES pH 7.6, 400 mM KCl, 10 mM MgCl₂, 2 mM EGTA, 10 mM 2-mercaptoethanol, 1 mM PMSF and 10% glycerol). 0.5 ml fractions were collected manually and analyzed by SDS-PAGE. Figure 5.9 shows that some of the impurities were removed as well as most degradation products. The amount of purified protein was estimated roughly from the Coomassie-stained band to be 200 µg. The accurate protein concentration measurement could not be performed since the sample contained DNA.

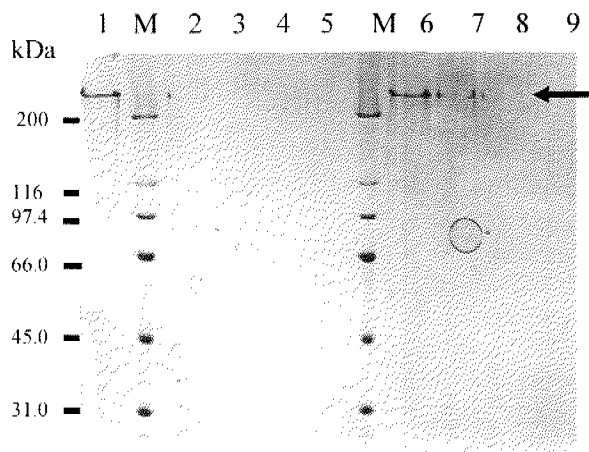


Figure 5.9. Calmodulin affinity purification with TALON™ elution fractions of hTAF1. Lane 1- column input, 2- unbound fraction, 3 to 5- wash fractions, 6 to 9- elution fractions. The arrow points to the position of the hTAF1 band. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

To analyze the DNA contaminants present in all hTAF1 purification fractions, samples were loaded on the 1% agarose gels and stained with EtBr. Many strategies were tried in order to remove the contaminating DNA. The attempt to soak hTAF1 out of the purified Sf21 nuclei was not successful. DNA from the nuclear extract could be removed using the hydroxyapatite resin, but not completely: the rest of DNA was co-eluting with hTAF1 if the hydroxyapatite flow-through was loaded on the Resource Q column. The hydroxyapatite purification step was done using the nuclear extract prepared as described above and incubated in batch with 4 ml of the resin for 45-60 min at 4° C. Figure 5.10 shows the results of Resource Q purification with the hydroxyapatite flow-through: hTAF1 was eluting in two main peaks, one together with most other nuclear proteins and a small amount of short DNA (100-200 bp), and the second one with higher purity in terms of protein contaminants but containing a large amount of DNA (500-1000 bp). If the nuclear extract is treated with 42 U/ ml MNase prior to the hydroxyapatite treatment, the DNA (mainly around 150 bp in size) is still co-eluting with hTAF1 from Resource Q but now in a single chromatographic peak. Also, when the hydroxyapatite flow-through was treated with 105 U/ ml of DNase I, DNA was degraded and hTAF1 precipitated. The attempt to remove the DNA contaminants with a 1 M KCl wash after loading on the TALON™ column was also not successful. Taken together, these experiments indicate that hTAF1 was present in high molecular weight aggregates together with DNA.

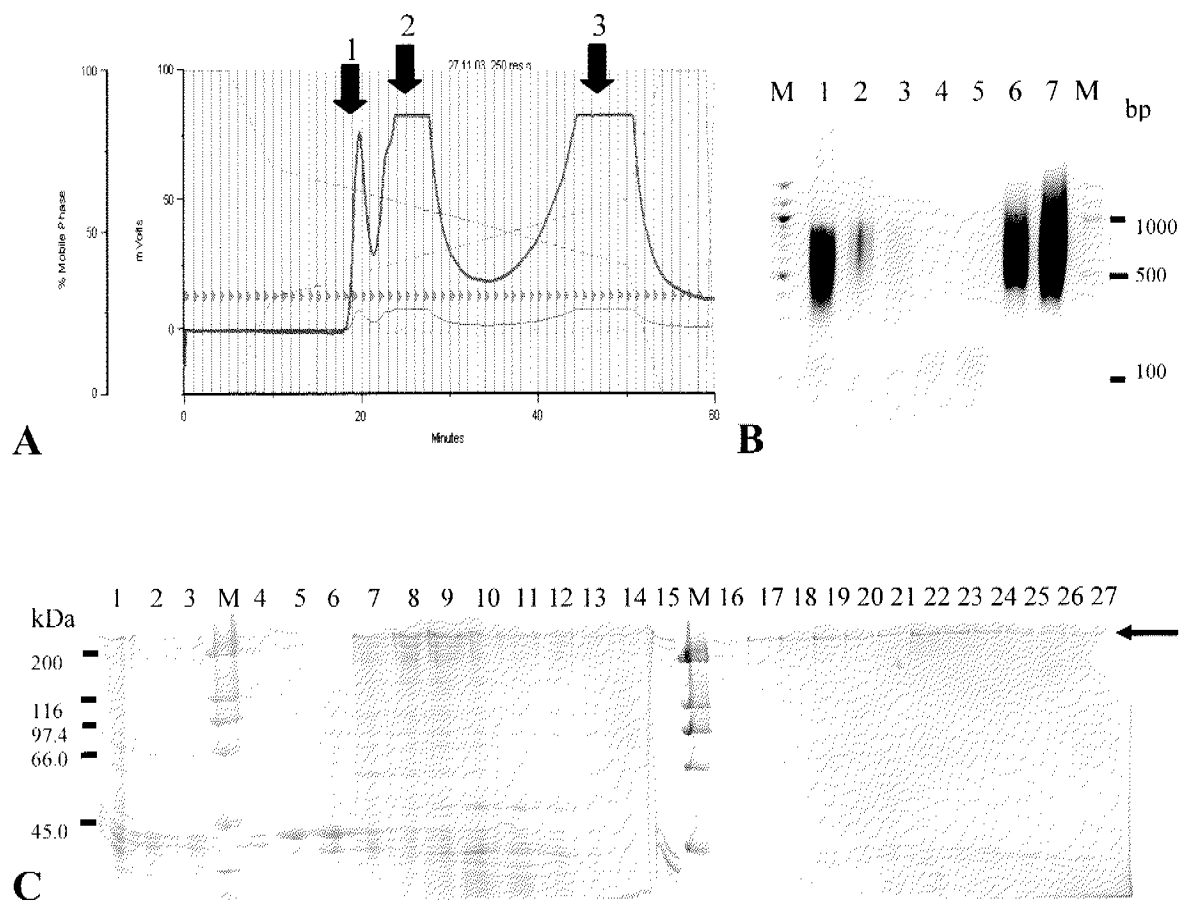


Figure 5.10. Analysis of DNA contaminants in hTAF1 purification fractions.

A) Resource Q purification chromatogram. Two main peaks are indicated by arrows (#2 and #3).
 B) Agarose gel analysis of the DNA content from two main peak fractions. Lane 1- nuclear extract, 2- sample after hydroxyapatite purification, 3- Resource Q input, 4 and 5- peak 2 fractions (#25 and 28), 6 and 7- peak 3 fractions (#45 and 50). M is the 100 bp DNA marker.
 C) SDS-PAGE analysis of the protein content from peak fractions. Lane 1- nuclear extract, 2- sample after hydroxyapatite purification, 3 and 15- Resource Q input, 4 to 6- peak 1 fractions, 7 to 14- peak 2 fractions, 16 to 27- peak 3 fractions. The arrow indicates the position of hTAF1 band. Lane "M" contains the broad range SDS-PAGE protein marker (BioRad).

5.2.2 Production of truncated hTAF1 constructs and their co-expression with hTAF2

Since purification of the full-length hTAF1 did not give material of acceptable quality and in sufficient quantities necessary for hTAF1-hTAF2 complex reconstitution, a co-expression strategy was attempted in order to produce the complex. Three truncation constructs of hTAF1 were designed for this purpose (Figure 5.11). Co-expression and co-purification of these constructs with the full-length hTAF2 would enable finding the domain of hTAF1 responsible for the interaction with hTAF2.

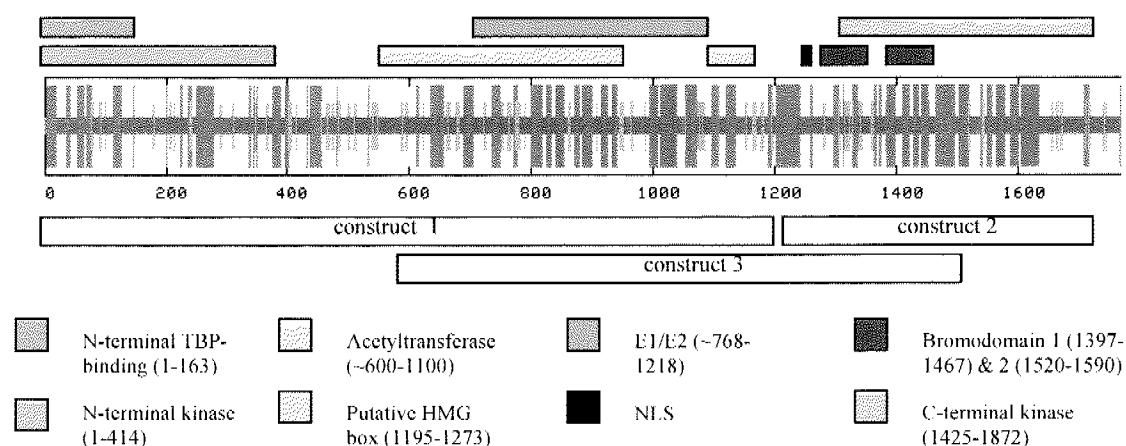


Figure 5.11. Design of hTAF1 domain expression constructs.

hTAF1 sequence (aa 1-1872) with the secondary structure prediction (α helix – blue, β strand – red, random coil – violet) is represented as a horizontal bar. Functional domains known from the literature are indicated above, with the color code below the figure. Designed hTAF1 constructs (1, 2, 3) are shown below the hTAF2 bar.

The three hTAF1 constructs were designed using phylogenetic and bioinformatics data. Construct 1 was designed based on the alignment of the hTAF1 with the yeast (*S. cerevisiae*) homolog. As mentioned before, yeast has a shorter version of TAF1 (145 kDa) corresponding to the N-terminal 1300 amino acids of the higher metazoan TAF1, apparently without the E1/E2 domain. The missing piece, containing the bromodomains and the C-terminal kinase, is found to exist as a separate protein Bdf1 which interacts with yeast TFIID through TAF7 (Matangkasombut et al, 2000). Construct 2 was designed to correspond to Bdf1 in yeast, since there was high probability that it will be expressed as a soluble protein. Construct 3 was designed to partially overlap with the other two constructs, containing the phylogenetically most conserved hTAF1 region also rich in secondary structure.

Properties of three hTAF1 expression constructs are described in detail in the Table 5.2.

CONSTRUCT	DESCRIPTION
1- based on yeast TAF1 (TAFII145)	Amino acids: 1-1293 MW: 152.5 kDa NLS: no Domains: TBP binding, N-terminal kinase acetyltransferase E1/E2 HMG box
2- based on yeast Bromodomain factor 1	Amino acids: 1293-1872 (end) MW: 71.2 kDa NLS: yes Domains: Bromodomains C-terminal kinase
3- based on conserved and structured middle part of the protein	Amino acids: 630-1638 MW: 122.5 kDa NLS: yes Domains: acetyltransferase HMG box E1/E2 Bromodomains

Table 5.2. Properties of designed hTAF1 domain constructs.

5.2.2.1 Cloning of hTAF1 constructs

All three hTAF1 constructs were designed to have the CBP-STREP affinity tag at the N-terminus, previously constructed as a part of the full-length hTAF1 expression vector pFBDOpolhTAF1 (section 5.2.1.2). Together with the internal His-tag contained within hTAF2, this tag could be used for purification of the hTAF1-hTAF2 complex.

pFBDO-based expression vectors were constructed by a PCR reaction over the desired region of the hTAF1 gene, using appropriate primers containing *Nde* I (5'-end) and *Rsr* II (3'-end) restriction sites. Two STOP codons were introduced at the end of each sequence removing the vector encoded His-tag. PCR products were subcloned first into the TOPO vector, using the procedure described in the section 2.2.19. The *Nde* I – *Rsr* II fragment was subsequently transferred into the pFBDOpolhTAF1 expression vector, replacing the full-length hTAF1 gene. The identity of all constructs was confirmed by sequencing.

5.2.2.2 Expression of hTAF1 constructs in Sf21 insect cells

The recombinant viruses carrying short hTAF1 construct genes were produced according to the procedures described in the section 2.4. After the initial transfection, viruses were amplified twice before the expression was detected on the SDS-PAGE. All three constructs were expressed at very high levels, but only construct 2 was present in the soluble fraction (Figure 5.12). Since it was possible that small amount of soluble protein was not detected by SDS-PAGE, purification of the constructs 1 and 3 was attempted by incubating a large amount of the protein extract with Calmodulin affinity resin. Unfortunately, the proteins could not be detected in the soluble fraction. Additional attempts were made to recover some material by dissolving the insoluble fraction in 8 M urea. This step was followed by stepwise urea removal to achieve refolding in the presence or absence of purified full-length hTAF2. None of these procedures were successful in obtaining soluble hTAF1 construct 1 and construct 3, alone or in the complex with hTAF2.

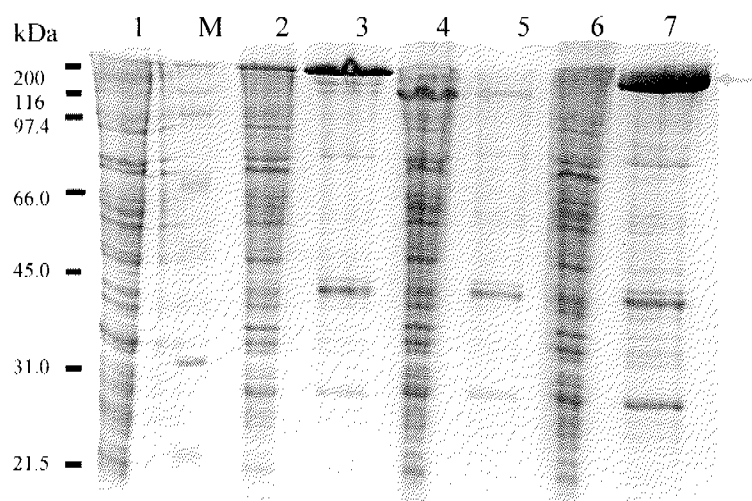


Figure 5.12. Expression and solubility test of hTAF1 domain constructs.

Lane 1- negative control, 2- construct 1 soluble fraction, 3- construct 1 insoluble fraction, 4- construct 2 soluble fraction, 5- construct 2 insoluble fraction, 6- construct 3 soluble fraction, 7- construct 3 insoluble fraction. The arrows are pointing to the respective protein bands. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

5.2.2.3 Purification of hTAF1 C-terminal domain (construct 2) and hTAF2 interaction analysis

hTAF1 construct 2 was expressed solubly at high level in Sf21 insect cells. The purification of this construct was done in batch using the Calmodulin affinity resin. The Sf21

cell pellet from the 200 ml of the expression culture was resuspended in 6 ml of the Binding buffer (25 mM HEPES pH 7.6, 400 mM KCl, 2 mM CaCl₂, 10 mM 2-mercaptoethanol, 1 mM PMSF and 0.1% NP-40). The cells were sonicated 2 x 10 s and centrifuged (12000 rpm, 10 min, 4° C, SS-34). The supernatant was rotated for 30 min at 4° C with 250 µl of the Calmodulin affinity resin pre-equilibrated in the same buffer. The unbound fraction was removed by centrifugation (2000 rpm, 2 min, 4° C, tabletop) and the resin was washed with 15 ml of the Binding buffer. Finally, the protein was eluted with 800 µl of the Elution buffer (25 mM HEPES pH 7.6, 400 mM KCl, 2 mM EGTA, 10 mM 2-mercaptoethanol, 1 mM PMSF and 0.1% NP-40). Figure 5.13 shows the results of the purification. hTAF1 construct 2 was successfully eluted from the resin at high purity. Interestingly, the protein band was traveling on the SDS gel higher than expected from its molecular mass (71.2 kDa).

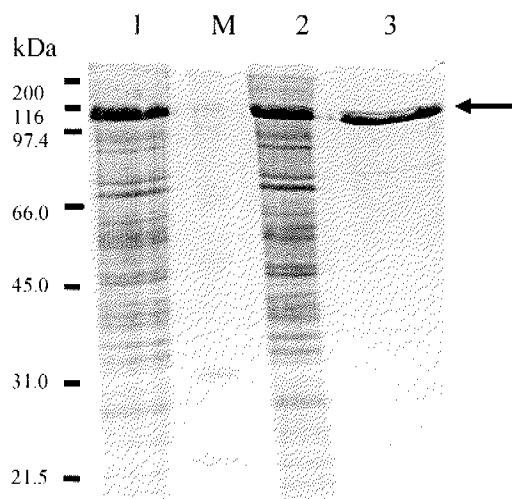


Figure 5.13. Purification of the soluble C-terminal hTAF1 domain (construct 2).

Lane 1- whole cell extract (soluble fraction), 2- Calmodulin affinity resin unbound fraction, 3- Calmodulin affinity resin elution fraction. The arrow points at the purified protein position. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

To test if hTAF1 construct 2 can form a complex with hTAF2, the above procedure was repeated in order to bind this construct to the Calmodulin affinity resin. Before elution with 2 mM EGTA-containing Elution buffer, the resin with the hTAF1 construct 2 was incubated with the purified full-length hTAF2 in the modified Binding buffer containing 200 mM KCl. The flow-through was collected, the resin extensively washed with 90 resin volumes of Binding buffer, and the sample eluted with the Elution buffer. The results showed that hTAF1 construct 2 does not form a complex with hTAF2, since the hTAF1 construct eluted without hTAF2 (Figure 5.14).

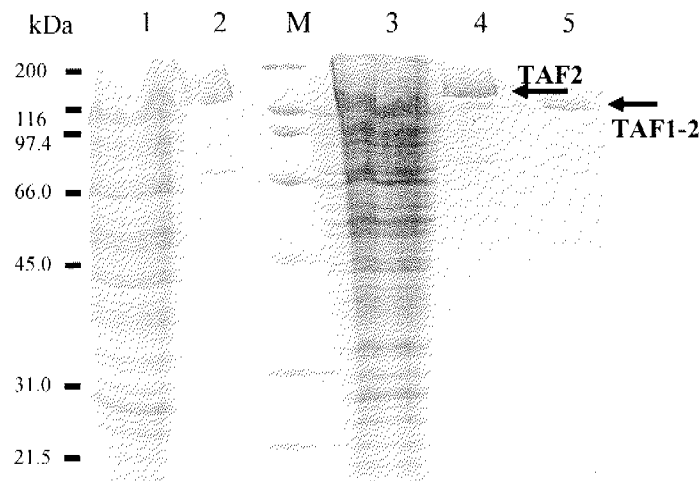


Figure 5.14. Test of hTAF1 construct 2 interaction with hTAF2.

Attempt of hTAF1 construct 2-hTAF2 complex reconstitution on the Calmodulin affinity resin. Lane 1- whole cell extract with hTAF1 construct 2, 2- purified hTAF2, 3- unbound fraction of hTAF1 construct 2 whole cell extract, 4- unbound fraction after incubation with hTAF2, 5- elution fraction.

The arrows point to the TAF1 construct 2 and TAF2 proteins. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

5.2.2.4 Co-expression of hTAF1 constructs 1 and 3 with hTAF2

Since hTAF1 constructs 1 and 3 were insolubly expressed on their own, an attempt was made to solubilize them by co-expression with hTAF2. 50 ml cultures of Sf21 cells at the density of 1×10^6 cells/ml were co-infected with the full-length hTAF2- and hTAF1 construct-expressing viruses. The amount of each virus used for infection was previously adjusted to give comparable expression levels of both proteins, since the expression of hTAF1 constructs was much higher than hTAF2. Figure 5.15 shows the results of this experiment: even if co-expressed with hTAF2, hTAF1 constructs 1 and 3 were found in the insoluble fraction. In case that some soluble material was present but undetectable by SDS-PAGE, an attempt was made to purify the complex using the Calmodulin affinity resin pull-down (procedure described in the section 5.2.2.3). The complex could not be detected in the elution fractions.

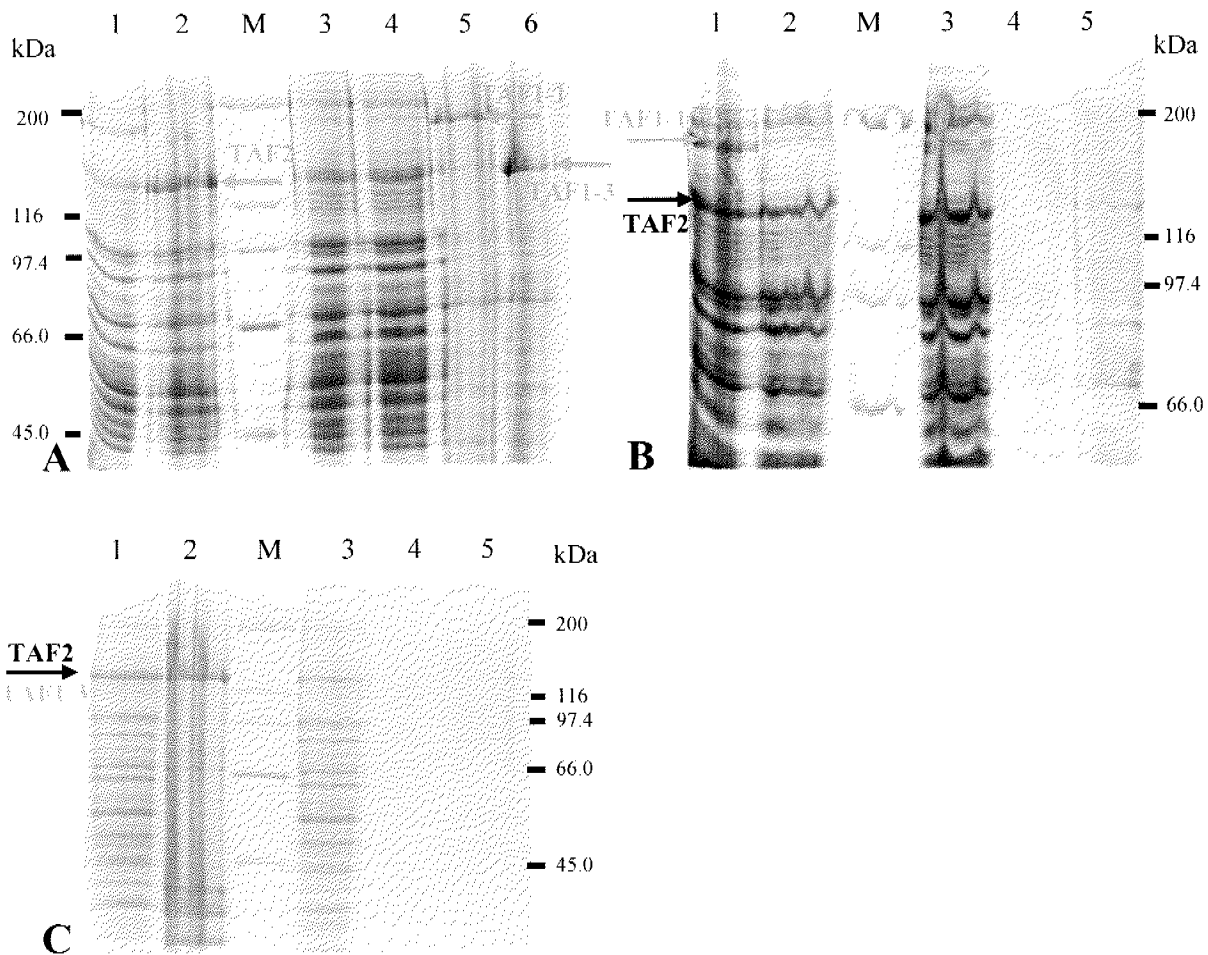


Figure 5.15. Co-infection of hTAF1 domain construct 1 and 3 viruses with the hTAF2 virus.

A) Soluble and insoluble fractions from the whole cell extract. Lanes 1 and 2- whole cell extracts of hTAF2 co-expression with hTAF1 construct 1 and 3, respectively; 3 and 4- respective soluble fractions, 5 and 6- respective insoluble fraction.

B) Calmodulin affinity resin pull-down with hTAF1 construct 1. Lane 1- whole cell extract, 2- soluble fraction from the whole cell extract, 3- unbound fraction, 4- elution fraction.

C) Calmodulin affinity resin pull-down with hTAF1 construct 3. Lane 1- whole cell extract, 2- insoluble fraction from the whole cell extract, 3- unbound fraction, 4- wash fraction, 5- elution fraction. Red arrows point to the position of hTAF1 construct bands. Black arrows point to the position of hTAF2. Lane "M" contains the broad range SDS-PAGE protein marker (BioRad).

5.2.3 Co-expression of hTAF1, hTAF2 and hTBPc in Sf21 insect cells

Production of the hTAF1-hTAF2 complex was also attempted in a co-expression construct containing full-length forms of these two proteins and hTBP core. This construct was cloned in a three-step procedure. First, the kanamycin resistance gene with its promoter region was PCR amplified from the pIBHO plasmid using MMKan5 and MMKan5 primers carrying the *BstE* II restriction site. The PCR product was ligated into the hTAF1 gene from the pFBDOpolhTAF1 vector using *BstE* II (Figure 5.16). This step was necessary to prevent

recombination events in the hTAF1 gene that made direct subcloning of the hTAF1 expression cassette technically impossible. In the second step, the resulting pFBDOpolh250Kan vector was cut with *Pme* I and *Spe* I and ligated with the insert containing hTAF2 and hTBPC genes previously released from the pDiFB150TBPcore vector (obtained from Dr. I. Berger) using *Avr* II and *Sna*B I restriction enzymes. This ligation was possible since *Avr* II and *Spe* I have compatible overhangs, while *Pme* I and *Sna*B I produce blunt ends. Positive colonies were selected using the ampicillin, gentamycin and kanamycin-containing agar plates. The third cloning step was the removal of the kanamycin resistance gene from the hTAF1 gene by *Bst*E II digest. The resulting final expression construct (pFBDOtriple) was used for the preparation of the recombinant baculovirus.

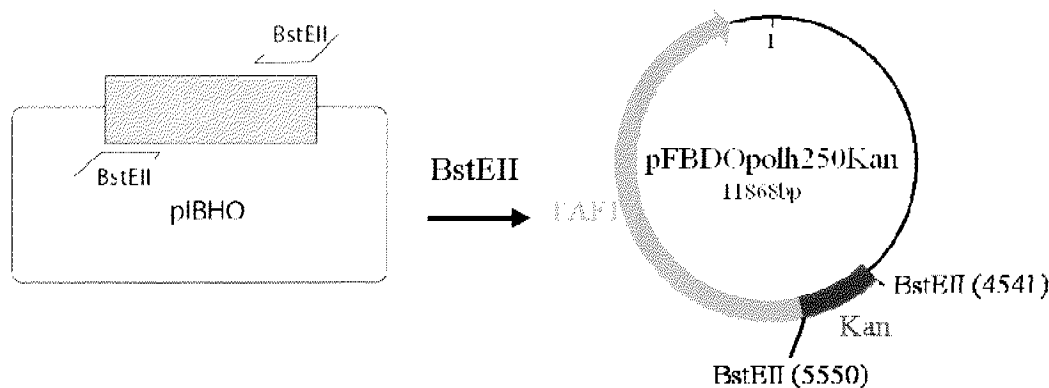


Figure 5.16. Cloning strategy for the hTAF1-hTAF2-hTBPC co-expression construct. Kanamycin resistance gene was transferred from the pIBHO plasmid to the beginning of the hTAF1 gene in pFBDO expression vector. This modified hTAF1 gene was cut out and used for the ligation into the hTAF1-hTAF2-hTBPC co-expression construct.

The recombinant virus carrying hTAF1, hTAF2 and hTBPC genes was produced according to the procedures described in section 2.4. After the initial transfection, the virus was amplified twice before the expression was analyzed by SDS-PAGE (Figure 5.17). hTAF2 was expressed from this construct at levels comparable to the construct containing only hTAF2, while hTAF1 expression could only be detected by Western blot analysis using the His-tag antibody (Figure 5.18). Expression of hTBP core was not detected by SDS-PAGE. One reason could be that the gene was not expressed, but it is very likely that the expressed band was not strong enough to be visible among many insect cell proteins in this molecular weight region. Purification of the complex from the whole cell extracts as well as

nuclear/cytosolic fractions was attempted using the TALON™ or Calmodulin affinity resin. The complex was not isolated even if large volumes of the protein extracts were used, indicating that it was either not formed at all, or it was present in minute amounts due to a low level of hTAF1 expression.

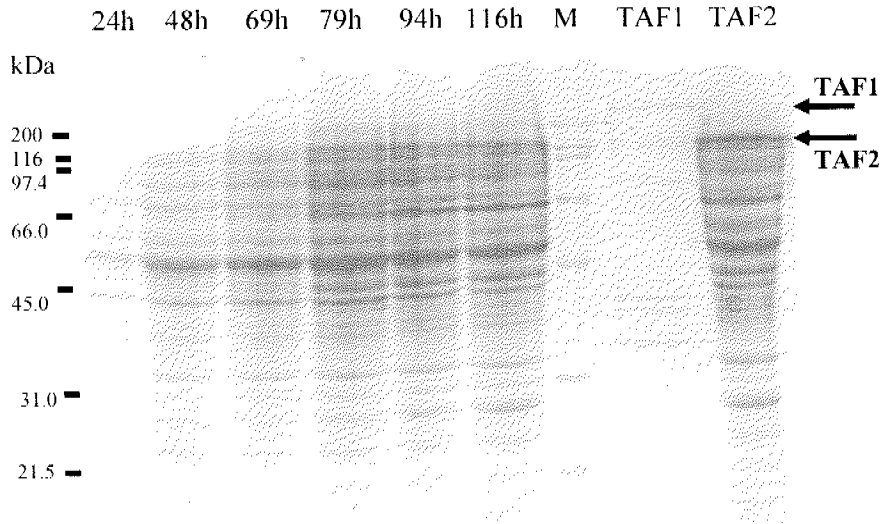


Figure 5.17. Expression time course from the hTAF1-hTAF2-hTBPC virus. Separate hTAF1 and hTAF2 samples are loaded as a positive control. The arrows point to the position of the bands. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

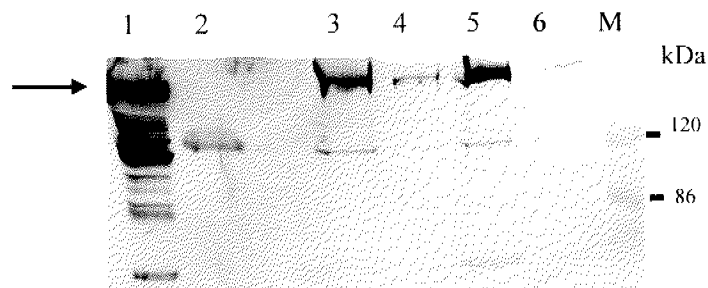


Figure 5.18. Detection of hTAF1 expression in the hTAF1-hTAF2-hTBPC virus by anti-His western blot.

Lane 1- purified hTAF1, 2- purified TAF2, 3 to 6- four different hTAF1-hTAF2-hTBPC viruses. The arrow points to the position of the hTAF1 signal. Lane “M” is the prestained protein marker (BioRad).

5.3 Interaction of hTAF2 with hTAF8/10

5.3.1 Interaction of full-length hTAF2 with hTAF8/10 complex

To investigate the interaction between hTAF2 and hTAF8/10 complex, two approaches were taken: gel filtration of the complex mixtures and Biacore analysis. TAF8/10 complex was expressed in Sf21 insect cells using the recombinant baculovirus obtained from Dr. I. Berger. The complex was purified according to the protocol described in the section 2.5.6. The final purification step was the gel filtration column (S200), where the complex eluted after 11.5 ml, indicating a tetrameric oligomerization state (the calculated molecular weight of hTAF8/TAF10 tetramer is 114.8 kDa). On the same S200 column, using the same buffer (25 mM HEPES-NaOH pH 7.0, 100 mM KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF), hTAF2 (137 kDa) elutes after 11 ml. To assay the interaction between hTAF2 and hTAF8/10 complex, hTAF2 was mixed with hTAF8/10 at 0.3 mg/ml concentration of both components, having hTAF8/10 in small excess (1:1.2 molar ratio of hTAF2 and hTAF8/10 tetramer). The mixture was incubated 30 min at room temperature, centrifuged (12000 rpm, 10 min, 4° C, benchtop) and loaded on the S200 column. The column was pre-equilibrated in the S200 buffer (25 mM HEPES-NaOH pH 7.0, 100 mM KCl, 10 mM MgCl₂, 1.5 mM DTT and 1 mM PMSF), which was used as a running buffer as well. There were no visible signs of precipitation in the protein mixture or any pellet after the centrifugation step. The 0.5 ml fractions were collected during the run, concentrated 10x in the Amicon concentrating device and analyzed by SDS-PAGE. Figure 5.19 shows the results of the experiment: one main peak was eluting after 7 ml, indicating that most of the sample was in high molecular weight aggregates. This peak contained both hTAF2 and hTAF8/10, apparently at 1:1 molar ratio judging from the SDS-PAGE. Interestingly, there was a small peak after 11.5 ml, which contained only the hTAF8/10 complex. Since the hTAF8/10 complex was in molar excess in this experiment, it was likely that hTAF2 indeed interacted with hTAF8/10. The interaction produced a complex that eluted at (or near) the void volume of the column, in a form of soluble aggregates.

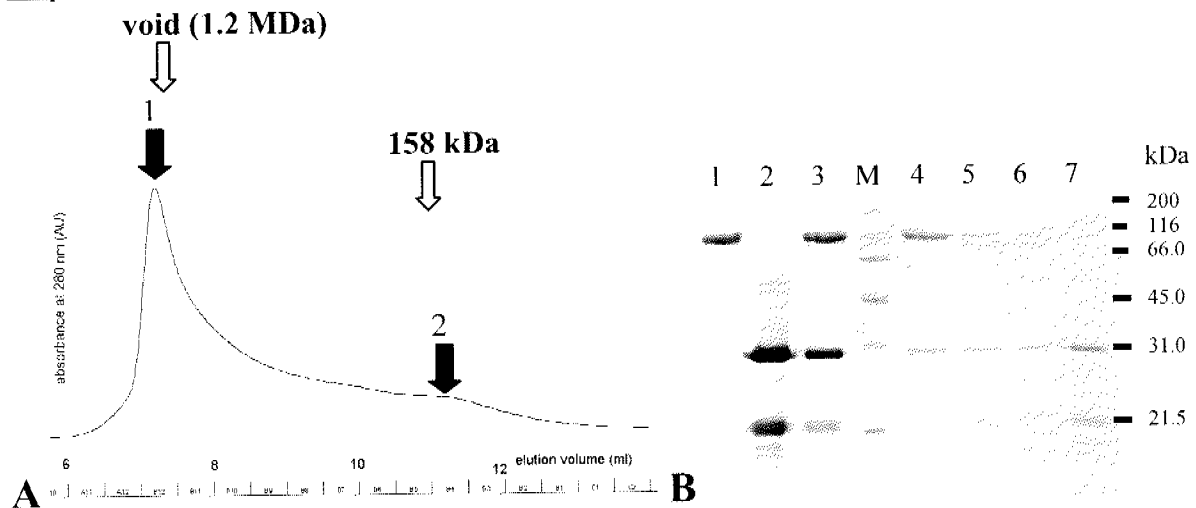


Figure 5.19. Attempt to detect the hTAF2-TAF8/10 complex by gel filtration (S200 column). A) Chromatogram, showing two main peaks: 1 (7-8 ml) and 2 (11 ml). B) SDS-PAGE of the fractions. Lane 1- hTAF2, 2- hTAF8/10 complex, 3- mixture of hTAF2 and hTAF8/10 (column input), 4- 8 ml fraction, 5- 10 ml fraction, 6- 11 ml fraction, 7- 12 ml fraction. Lane “M” contains the broad range SDS-PAGE protein marker (BioRad).

The Biacore system, described in detail in section 2.8, was used as another method to investigate the interaction between hTAF2 and hTAF8/10. The goal of these experiments was to get a qualitative answer whether or not hTAF2 and hTAF8/10 interact, and possibly to obtain dissociation/association and reaction rate constants.

A set of initial experiments was performed in order to find the best conditions for measurement. As a first step, hTAF2 was injected over different Biacore chips (CM5, CM4, CM1, SA) to assess the non-specific binding to the chip surface. Since this protein appeared to have a very strong non-specific affinity for the surface, it was decided to have it as the immobilized ligand rather than to use it as the analyte. hTAF8/10 also had some non-specific affinity for the chip surface, but it could be washed away easily with the running buffer. Therefore, hTAF8/10 was more suitable as the analyte in Biacore binding experiments.

The second step was the development of hTAF2 immobilization procedure. Since the CM4 chip allowed better control over the amount of the immobilized ligand (having less carboxyl groups for covalent coupling), as well as had less non-specific interaction with hTAF8/10, it was chosen for all the following experiments. hTAF2 could be immobilized using either amine or thiol coupling procedure, but amine coupling was chosen because of its speed and simplicity. For the immobilization reaction, hTAF2 was prepared in the buffer containing 25 mM HEPES-NaOH pH 7.6, 100 mM KCl, 1.5 mM DTT and 1 mM PMSF. The pH of this buffer allowed the protein to be concentrated near the chip surface for efficient immobilization, since the negatively charged surface attracted the positively charged protein.

Different immobilization levels (from several hundred to several thousand RU) were tried in the course of these experiments. The immobilization level did not have an influence on the results.

The next step was to inject different concentrations of hTAF8/10 over the chip surface and compare the response of the non-active versus active (hTAF2-containing) surface. The non-active surface (reference lane) was prepared using the same chemical procedure as the active surface, but omitting the hTAF2 injection step. If the response of the active surface was higher than the response of the non-active surface, it would indicate that interaction is indeed taking place. In this experiment, the answer was positive, having an order of magnitude higher response on the hTAF2-containing surface than in the non-active one. Additionally, as expected, the response was increasing with the increasing hTAF8/TAF10 concentration.

In order to obtain quantitative information on the hTAF2 interaction with hTAF8/10 complex, it was necessary to find the conditions that would allow chip surface regeneration (rapid removal of hTAF8/10 from the active surface). This was important in order to be able to inject many different hTAF8/TAF10 concentrations over the same chip. Multiple injections would allow the acquisition of data necessary to estimate the kinetic and affinity parameters of the reaction. After varying many regeneration buffer components (salt concentration, pH, detergents), the following buffer was found to be most effective, allowing high reproducibility of duplicate injections and baseline stability: 20 mM HEPES-NaOH pH 8.0, 160 mM KCl, 0.8 mM DTT, 0.8 mM PMSF, 0.008% Tween-20 and 0.05% SDS.

After the regeneration conditions were established, the Biacore run could be performed. Data of best quality was obtained with the running buffer containing 25 mM HEPES-NaOH pH 8.0, 200 mM KCl, 0.01% Tween-20, 1 mM PMSF and 1 mM DTT. After immobilizing 480 RU of the hTAF2 by stepwise injection at the concentration of 0.1 mg/ml, the following run was set up:

hTAF8/TAF10 (tetramer) concentrations:

44 nM, 87 nM, 218 nM, 436 nM, 653 nM, 1.1 μ M

Injection: 5 min

Dissociation: 5 min

Regeneration: 4 s injection at 80 μ l/min

Flow rate: 5 μ l/min

Stabilization time after regeneration: 10 min

Figure 5.20 A shows the resulting Biacore sensograms curves (difference between the active and non-active lanes). The qualitative shape of the curves indicates rapid association and slow dissociation. Rapid association and dissociation is seen usually in the interactions of proteins with small molecules. Rapid association of hTAF2 and hTAF8/10 could be explained by a non-specific binding as a first step, followed by formation of the specific complex. Since the association phase is so rapid, the curves could only be used for the steady state binding analysis, taking the equilibrium response levels at different hTAF8/TAF10 concentrations for the K_d calculation by BiaEvaluation software. The model used for data fitting was 1:1 binding, and the analysis was done according to the Biacore user manual. The resulting fit is shown on the Figure 5.20 B. Maximum hTAF2 response (R_{max}) was extrapolated to be 366 RU, and the Chi^2 of 55.7 indicated the quality of the fit. The lowest hTAF8/TAF10 concentration (44 nM) was omitted in order to obtain a better fit. Generally, the fit is considered to be good if the Chi^2 is an order of magnitude smaller than the R_{max} (BIAtechnology handbook, 1994). The reason for the bad fit of 44 nM hTAF8/10 with the rest of the data could be the possibility that the hTAF8/10 complex exists in a dimer-tetramer equilibrium, dissociating to dimers at lower hTAF8/10 concentrations. The data for concentrations hTAF8/10 lower than 44 nM indicated the same. The K_d of the hTAF2-hTAF8/10 complex was calculated to be 101 nM, a value in the expected range for most specific protein-protein interactions. However, this is just a tentative value since a better fit would have to be obtained in order to get the accurate K_d .

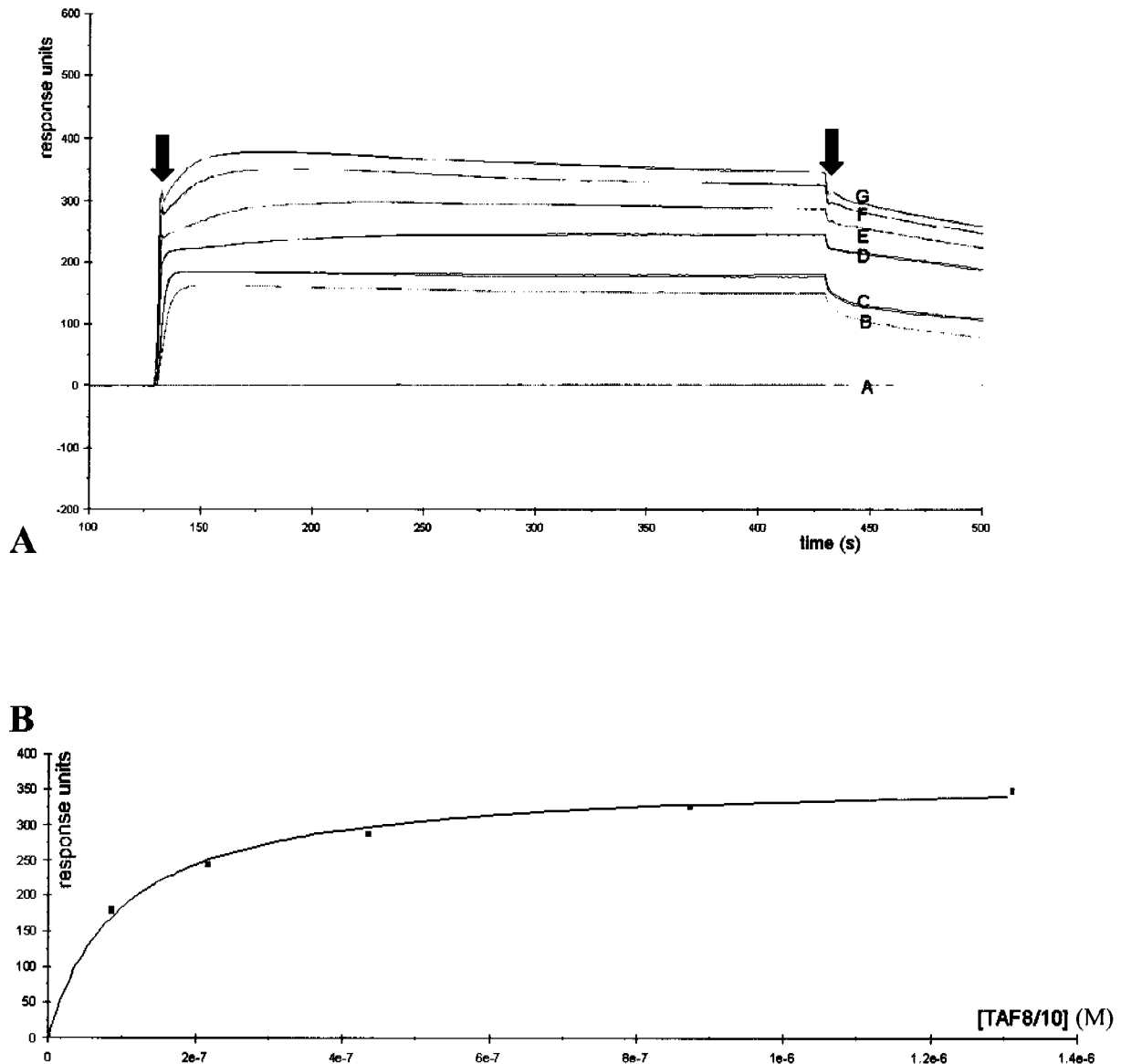


Figure 5.20. Binding of hTAF8/10 to the full-length hTAF2.

A) Biacore sensogram curves. Reference lanes are subtracted from the active lanes and corrected by the zero hTAF8/TAF10 concentration. Two arrows point to the start and end of the injection phase. hTAF8/TAF10 concentrations (all injected in duplicate): 0, 44 nM, 87 nM, 218 nM, 436 nM, 653 nM, 1.1 μ M (curves A to G, respectively). All hTAF8/10 concentrations are calculated for the tetrameric oligomerization state.

B) Data analysis. X-axis shows the analyte concentration and the Y-axis the corresponding signal amplitude after attainment of equilibrium. Experimental data (points) are fitted by the BiaEvaluation software according to the 1:1 binding model (steady state analysis). The 44 nM hTAF8/TAF10 concentration is omitted from the fit.

5.3.2 hTAF8/10 binding activity of hTAF2 domains

After establishing hTAF8/10 binding activity of the full-length hTAF2, hTAF2 core and tail domains were tested in order to find the domain responsible for this activity. hTAF2 core domain was immobilized on the CM4 chip by a stepwise injection at the concentration

of 0.5 mg/ml. The immobilization buffer contained 50 mM MES-HCl pH 6.0, 150 mM KCl, 1.5 mM DTT and 1 mM PMSF. The final immobilization level was 402 RU. Since the initial binding tests showed that hTAF2 core has hTAF8/10 binding activity, a run was performed using the following running buffer: 25 mM HEPES-NaOH pH 8.0, 150 mM KCl, 0.1% Tween-20, 1 mM PMSF and 1 mM DTT. Unfortunately, the running and regeneration buffers established for the full-length hTAF2 did not give the best results with the core domain, so their compositions had to be further adjusted. The regeneration buffer that showed the best results was: 20 mM HEPES-NaOH pH 8.0, 120 mM KCl, 0.8 mM DTT, 0.8 mM PMSF, 0.08% Tween-20 and 0.1% SDS. Run setup was the following:

hTAF8/TAF10 (tetramer) concentrations:

87 nM, 175 nM, 350 nM, 700 nM, 1.4 μ M, 2.8 μ M

Injection: 5 min

Dissociation: 5 min

Regeneration: 4 s injection at 80 μ l/min

Flow rate: 20 μ l/min

Stabilization time after regeneration: 10 min

Figure 5.21 A shows the resulting Biacore sensograms curves (difference between the active and non-active lanes). Qualitative shape of the curves is similar to the full-length hTAF2 curves, showing rapid association and slow dissociation. The data was analyzed as described for the full-length hTAF2, omitting the highest hTAF8/TAF10 concentration (2.8 μ M). The resulting fit is shown on the Figure 5.21 B. Maximum ligand response (R_{max}) was extrapolated to be 135 RU, with Chi^2 value of 83.3. The K_d of the hTAF2-hTAF8/10 complex was calculated to be 83 nM. Since the Chi^2 value was so high, the calculated K_d value can not be taken as reliable. Additional fine adjustment of running buffer composition could provide data of sufficient quality to get the accurate K_d value. Nevertheless, this experiment provided valuable information that hTAF2 core also had hTAF8/10 binding activity.

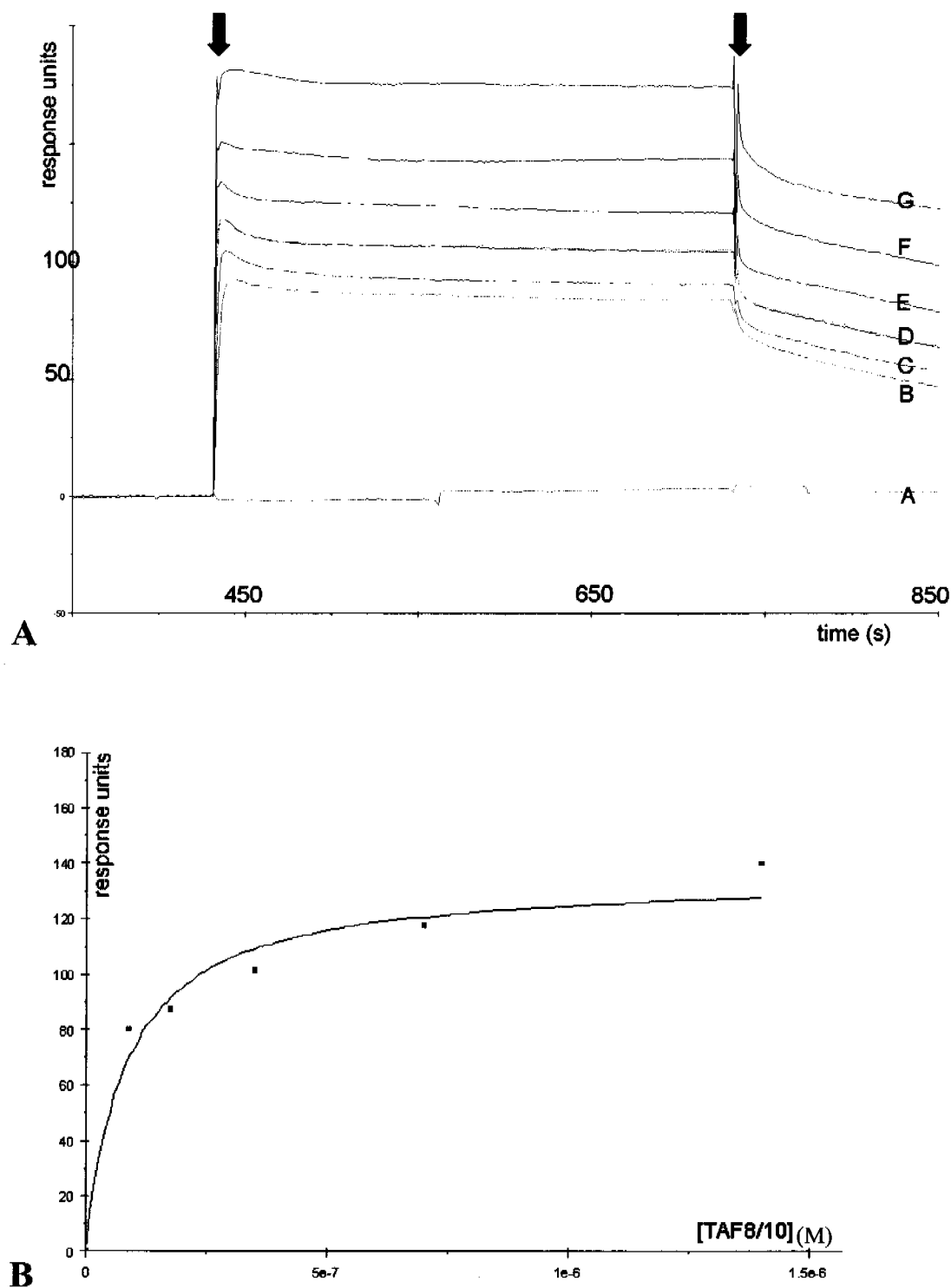


Figure 5.21. Binding of hTAF8/10 to the hTAF2 core.

A) Biacore sensogram curves. Reference lanes are subtracted from the active lanes. Two arrows point to the start and end of the injection phase. hTAF8/TAF10 concentrations: 0 nM, 87 nM, 175 nM, 350 nM (duplicate), 700 nM, 1.4 μ M, 2.8 μ M (curves A to G, respectively). All hTAF8/10 concentrations are calculated for the tetrameric oligomerization state.

B) Data analysis. X-axis shows the hTAF8/TAF10 concentration and the Y-axis the corresponding signal amplitude after attainment of equilibrium. Experimental data (points) are fitted by the BiaEvaluation software according to the 1:1 binding model (steady state analysis). The 2.8 μ M hTAF8/TAF10 concentration is omitted from the fit.

To investigate its hTAF8/10 binding activity, hTAF2 tail domain was immobilized on the CM4 chip by a stepwise injection at the concentration of 0.25 mg/ml. The immobilization buffer contained 25 mM HEPES-NaOH pH 7.6, 200 mM KCl, 1.5 mM DTT and 1 mM PMSF. The final immobilization level was 337 RU. The binding tests were performed in the following running buffer: 25 mM HEPES-NaOH pH 8.0, 200 mM KCl, 0.01% Tween-20, 1.5 mM DTT and 1 mM PMSF. Several injections of hTAF8/10 at 420 nM concentration were performed, producing the response difference between the active and inactive lane of almost zero. Therefore, hTAF2 tail domain did not appear to have any hTAF8/10 binding activity.

5.4 hTAF2 DNA binding activity

5.4.1 Analysis of full-length hTAF2 DNA binding activity

DNA band shift experiments with hTAF2 were conducted as described in the section 2.6.7. The goal was to confirm the data from the literature on the DNA binding activity of this protein, as well as to analyze it further in terms of binding specificity and affinity. Different DNAs that were used for these experiments are listed in Table 5.3.

NAME	SEQUENCE (size)	DESCRIPTION
KAUFF	5'-ACATCAGAGCCCTCATTCTGGAGA-3' (24 bp)	Initiator consensus (Py Py A ₋₁ N T/A Py Py) underlined. (Kauffmann et al, 1996.)
GAG	5'-ACCC <u>CCGAGCCA</u> AAGGGGGA-3' (19 bp)	“GAG core sequence” underlined. (Martin et al, 1999.)
CYC	5'- AGGGAGCAGTGC GGGGTTTAAATCTGAG GCTAGGCTGGCTCTTCTCGG CG -3' (50 bp)	Human cyclin B1 core promoter, containing two “GAG” sequences and a non-canonical TATA-box. (Martin et al, 1999.)
CORE	5'-GATCGA GGGCGCC C TATAAA GACAGTTTGCAGAGCAGTG CCCTCATTCTC GATCGATCGAT CGAACGGAACG GACGTG GATCGATCGA-3' (88 bp)	Synthetic core promoter with BRE, TATA, initiator, MTE, DPE elements.
MarD	5'-ACCC <u>CCGAGCCA</u> AAGGGGGA-3' (19 bp)	DNA used in Martin et al, 1999.
MarDNEG	5'- ACCCGCTTAGCAGGGGGGA-3' (19 bp)	DNA based on MarD with mutated specific recognition sequence (for the negative control).

Table 5.3. DNA used in hTAF2 DNA binding experiments.

Figure 5.22 shows the results of the initial DNA band shift experiments, using ethidium bromide (EtBr) stained and radioactively labeled CYC DNA for comparison. For the EtBr gel (Figure 5.23 A), purified 4 μ M hTAF2 was incubated with different

concentrations of the DNA (0.5, 1.0 and 2.0 μM) in the buffer containing 10 mM HEPES-NaOH pH 7.5, 50 mM KCl, 10 mM MgCl_2 and 2% glycerol. The samples were loaded on a 10% polyacrylamide gel (BioRad system) pre-run for 1 h in 1xTBE buffer. The electrophoresis was conducted for 1.5 h at 80 V. The results clearly showed that hTAF2 has DNA binding activity. Protein-DNA complex that was formed did not migrate into the gel, but stayed in the well. This indicates formation of very high molecular weight protein-DNA complexes too large to enter the gel. To prevent this aggregation, which could have been caused by too high a protein concentration, this experiment was repeated using radioactively labeled CYC DNA (Figure 5.23 B). Using this strategy, the protein-DNA complex could in principle be detected even with much lower protein concentrations. Purified hTAF2 was incubated at different concentrations (0.04, 0.18, 0.4 and 1.8 μM) with 1 μM DNA in the buffer containing 25 mM HEPES-NaOH pH 7.6, 50 mM KCl, 10 mM MgCl_2 , 0.1 mM PMSF and 1.5 mM DTT. The samples were loaded on a 10% polyacrylamide gel (BioRad system) pre-ran for 1 h in 1xTBE buffer. The electrophoresis was conducted for 1.5 h at 80 V. The results confirmed previous findings – the protein-DNA complex was still aggregated, disabling quantitative studies of binding affinity. In order to find conditions for a 1:1 protein-DNA complex, these experiments were repeated extensively with a range of incubation buffers (various pH and salt conditions, as well as other additives), running buffers, DNA and protein concentrations, different DNAs, gel systems, gel percentages and incubation temperatures. All experiments gave the same results: the DNA-protein complex was formed, but did not migrate into the gel. The concentrations used also indicated weak binding (dissociation constant in μM range).

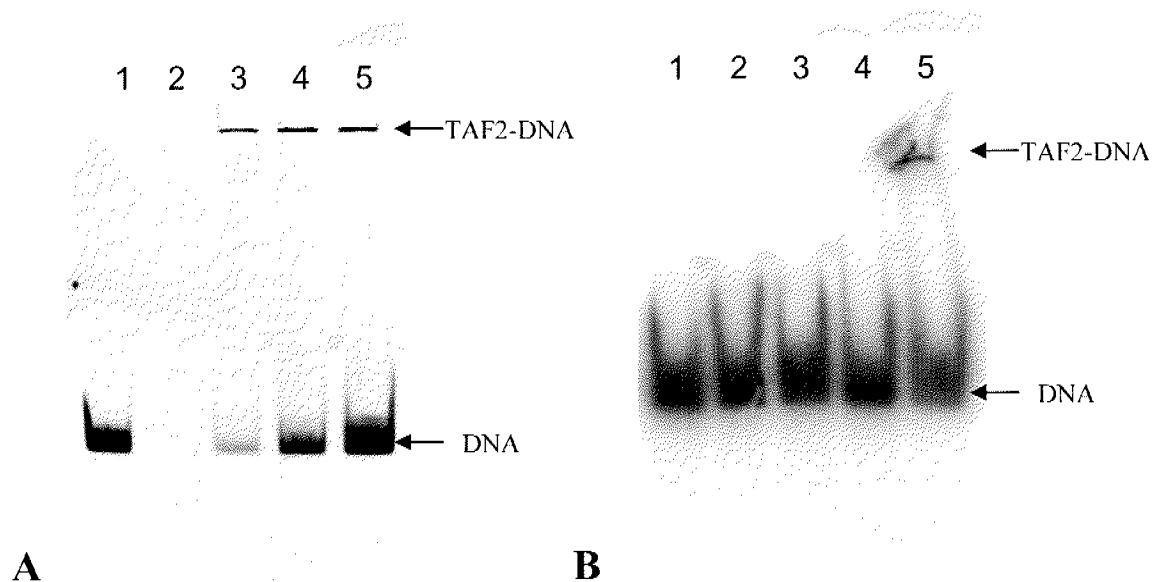


Figure 5.22. DNA-binding activity of hTAF2.

A) EtBr stained gel. Lanes 2-5 contain 4 μM hTAF2. Lane 1- 1.0 μM CYC DNA only, 3- 0.5 μM DNA, 4- 1.0 μM DNA, 5- 2.0 μM DNA. B) Radioactively labeled gel. All lanes contain 1.0 μM CYC DNA. Lane 1- no protein, lanes 2 to 5 contain 0.04, 0.18, 0.4 and 1.8 μM hTAF2 respectively. All samples had total volume of 25 μl .

Aggregation of the hTAF2-DNA complexes could be prevented only if the sample was loaded on the gel in a 0.05% (17 μM) SDS-containing loading dye. Non-ionic detergents (NP-40, Tween 20, Triton), as well as lower SDS concentrations were assayed to find the conditions that prevent hTAF2-DNA complex aggregation, but only 0.05% SDS was effective. Figure 5.23 shows the results of those experiments, done at the constant radioactively labeled GAG DNA concentration (12.5 nM) and various hTAF2 concentrations. For hTAF2 bandshift experiments, the samples were incubated in the GLc buffer (10 mM Tris-HCl pH 7.9, 10 mM HEPES-NaOH pH 7.9, 10% glycerol, 1 mM DTT, 4 mM MgCl_2 , 50 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 0.1 mg/ml BSA and 1.25 ng/ μl dl-cC competitor DNA) (Martin et al., 1999). The SDS-containing loading dye was added immediately before the samples were loaded on the gel. The electrophoresis was conducted in the TBM running buffer (89 mM Tris-HCl pH 7.5, 89 mM boric acid and 0.5 mM MgCl_2). Gels were ran at 100 V, with a pre-run for 2 h at 200 V. The voltage was initially adjusted to 200 V for 5 min to allow samples to enter the gel, and afterwards to 100 V for 4 h. Shifted bands obtained in these experiments were not clearly discrete, showing a smear that indicated a non-stable complex (smear is below the band) or a mixture of different complexes (no discrete band). At lower protein concentrations, the bands shifted to lower positions, followed by higher bands

at higher protein concentrations. This result indicated formation of a series of complexes with different stoichiometry. In all performed experiments, micromolar hTAF2 concentrations were necessary to detect the complex.

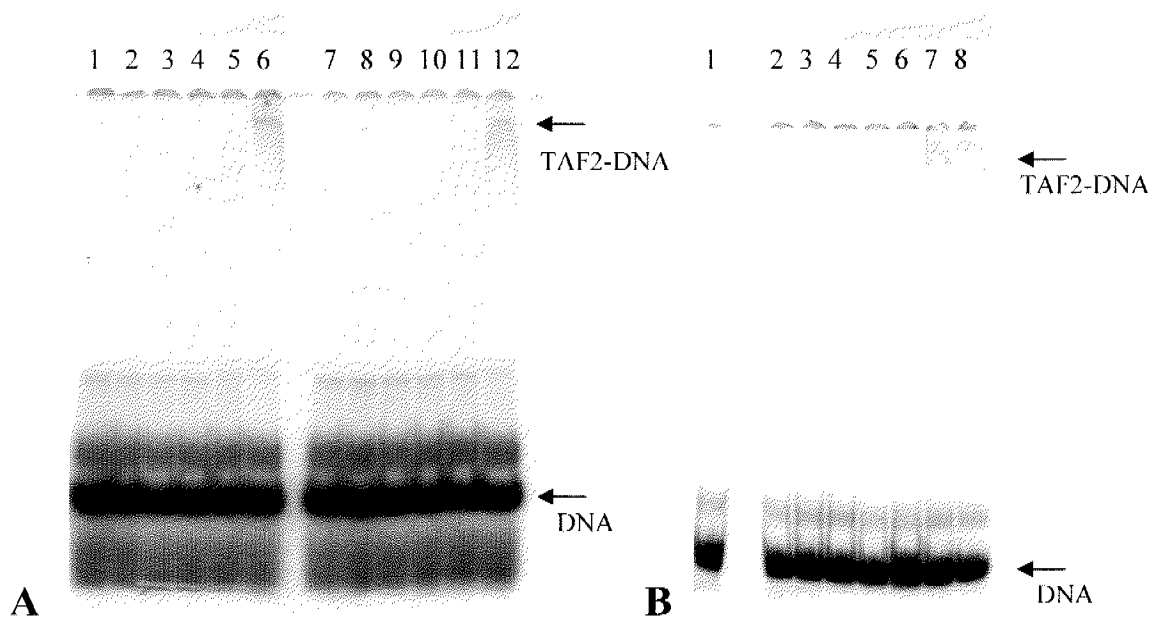


Figure 5.23. hTAF2 DNA binding assays with 0.05% SDS-containing loading dye.

A) Two-layer polyacrylamide gel (8% bottom layer, 5% top layer). All lanes contain 12.5 nM hot GAG DNA. Lanes 1 to 6 contain 1 μ M, and 7 to 12 contain 3 μ M cold DNA. Lanes 1 and 7 contain DNA only. Lanes 2 to 6 as well as 8 to 12 contain 17 nM, 52 nM, 170 nM, 510 nM and 1.7 μ M hTAF2.

B) 4% polyacrylamide gel. All lanes contain 12.5 nM hot GAG DNA. Lanes 2 to 8 contain 27 nM, 55 nM, 110 nM, 220 nM, 435 nM, 870 nM and 1.75 μ M hTAF2 respectively.

5.4.2 DNA binding activity of hTAF2 domains

In order to find the part of hTAF2 responsible for DNA binding, hTAF2 core and tail domains had to be analyzed separately. The experiment was done using radiolabeled GAG DNA (12.5 nM) and 0.35 to 3.5 μ M concentration of purified full-length hTAF2, hTAF2 core domain or hTAF2 tail domain. Samples were incubated in the GLc buffer and loaded on the gel with 0.05% SDS-containing loading dye (as described for the full-length protein, Figure 5.24). The electrophoresis was done in TBM running buffer, at 100 V for 4 h, with an initial 5 min pulse at 200 V. The results clearly show that the DNA-binding domain of hTAF2 is the tail domain, while the hTAF2 core domain has no DNA-binding activity on its own (Figure 5.24). Interestingly, the tail domain showed similar behavior as the full-length

protein, in terms of aggregation and the requirement to have at least 0.05% SDS in the sample for the complex to enter the gel.

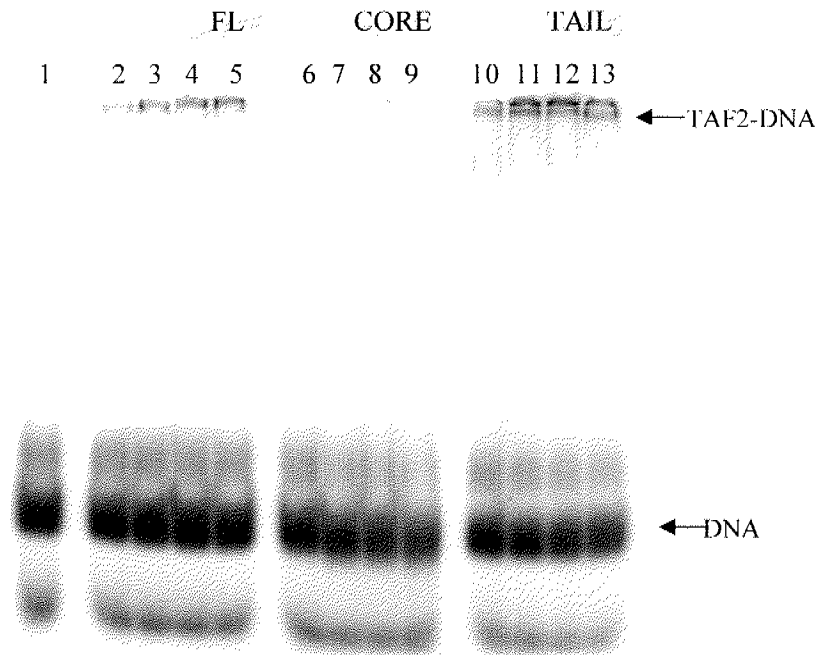


Figure 5.24. DNA binding activity of full-length hTAF2 and hTAF2 domains.

All lanes contain 12.5 nM radioactively labeled GAG DNA. Lanes 2-5 contain the full-length hTAF2, lanes 6-9 hTAF2 core domain and lanes 10-13 hTAF tail domain in 0.35, 0.70, 1.75 and 3.5 μ M concentrations.

To examine the specificity of hTAF2 DNA binding activity, full-length hTAF2 as well as the hTAF2 tail domain together with GAG DNA were incubated with different concentrations of cold dI-dC competitor DNA. Figure 5.25 shows the results of this experiment, where 1.75 μ M full-length hTAF2 or hTAF2 tail was mixed with 12.5 nM radioactively labeled GAG DNA (0.34 ng/ μ l) and 1.25 ng/ μ l to 125 ng/ μ l dI-dC DNA. At the highest competitor DNA concentration, which is 1000-fold larger than the concentration of radioactive DNA, the shifted band disappeared. The fact that the shift was still visible in the presence of 12.5 ng/ μ l competitor DNA (37x more than labeled DNA) indicates that hTAF2 has some sequence preference towards the GAG DNA as compared to the non-specific DNA.

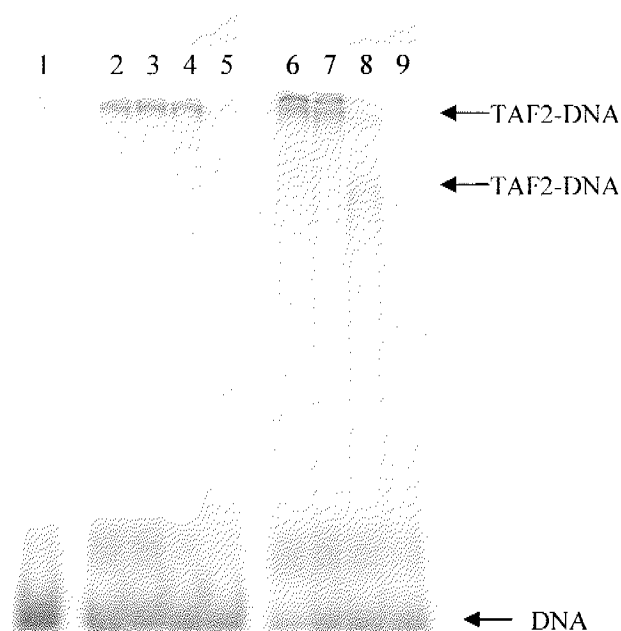


Figure 5.25. hTAF2 DNA binding tests with dI-dC competitor DNA.

All lanes contain 12.5 nM radioactively labeled GAG DNA (0.34 ng/ μ l). Lanes 2-5 contain the 1.75 μ M full-length hTAF2 and lanes 6-9 1.75 μ M hTAF2 tail. Lanes 2 and 6 have no competitor DNA, lanes 3 and 7 have 1.25 ng/ μ l dI-dC DNA, lanes 4 and 8 have 12.5 ng/ μ l dI-dC DNA and lanes 5 and 9 have 125 ng/ μ l dI-dC DNA.

CD spectrum of the hTAF tail domain showed a curve characteristic for unstructured proteins with a possibility of some α -helical content (section 4.4.3). To investigate if any structural changes of the tail domain occur upon DNA binding, this domain was mixed at 12.6 μ M (0.27 mg/ml) concentration with 1 μ M MarD or MarDNEG DNA (Table 5.3). After 30 min incubation at room temperature in the CD buffer, the spectra of the protein-DNA complexes were measured in the CD spectrometer (section 2.7.2). Figure 5.26 shows the resulting spectra obtained after subtracting respective DNA-only absorbance data. Compared to the protein-only spectrum, both DNA containing samples showed an increased signal at 222 nm. This wavelength is the position of one of two characteristic α -helical minima, the other being at 209 nm, a region “masked” by the strong unstructured signal at 200 nm. Sample containing MarDNEG DNA also showed a significant decrease in the unstructured signal. Taken together, these spectra indicate that structural changes do occur in the hTAF2 tail upon DNA binding, most likely increasing the α -helical content. A larger change in the spectra could be observed with at least 1:1 molar ratio of protein to DNA. However, this experiment was not possible to perform since it would require DNA stock solution at minimum 50-100 μ M concentration.

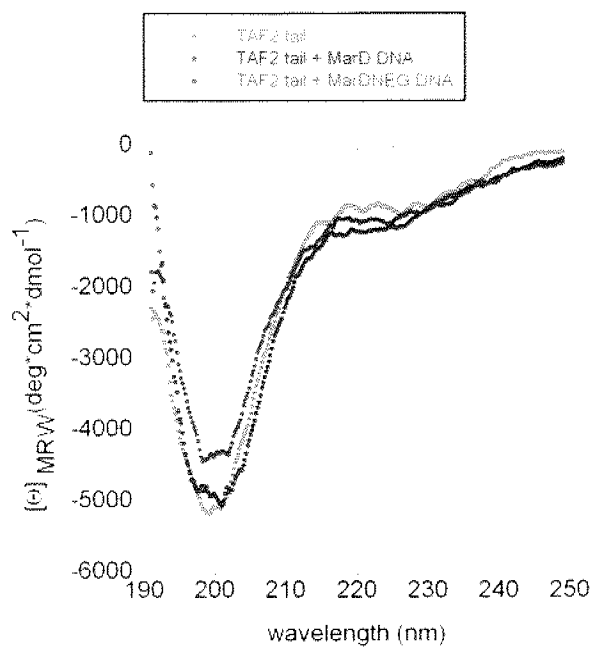


Figure 5.26. CD spectrum of hTAF2 tail domain with DNA.

5.5 Discussion and perspectives

Results presented in this chapter provided several insights into the function of human TAF2. Most importantly, hTAF2 was shown to interact directly with the hTAF8/10 complex, probably through its core domain. The tail domain was, on the other hand, shown to incorporate the DNA binding activity of hTAF2. This activity appeared largely non-specific, with the dissociation constant estimated to be in the micromolar range.

Extensive efforts were given to the hTAF1-hTAF2 complex production using hTAF1 either in full-length or its truncated versions, alone or in co-expression experiments with the full-length hTAF2. However, none of these strategies were successful in obtaining the complex. The existence of this TFIID subcomplex was never shown *in vivo*, but observed on successful reconstitution of hTAF1-dTAF2-d/hTBP complex from purified recombinant components *in vitro* (Chen et al, 1996). This complex was reported to support NTF-1 activated transcription in *in vitro* transcription assays (Chen et al, 1994). The same reports proposed that the TAF1 subunit is the scaffold of TFIID, a structural core upon which other TAFs are assembled. However, results presented in this chapter regarding hTAF1 purification show that this protein is highly unstable, as it is sensitive to proteolytic degradation and aggregation. When expressed in Sf21 insect cells, it is found in the nuclei strongly bound to chromatin in high molecular weight protein assemblies. If separated from DNA by DNase I digest, hTAF1 is lost from the soluble fraction, indicating that it loses its interaction partners that keep it soluble. Theoretical analysis of hTAF1 secondary structure indicates a protein that is not well folded on its own, since it has many large unstructured regions. Truncated versions of hTAF1, that were expressed insoluble on their own, were co-expressed with the full-length hTAF2 in an attempt to produce them in a soluble form. Since hTAF2, assumed to directly interact with hTAF1, was clearly not sufficient for its solubilization, there is a significant possibility that other TFIID subunits are necessary for structural stability of hTAF1. The interaction test of the C-terminal soluble domain of hTAF1 (construct 2) with hTAF2 was negative, implying that hTAF2 interaction domain is contained within the N-terminal 1293 amino acids of hTAF1. This result is not unexpected, since the yeast TAF1 homolog is a shorter protein without the corresponding C-terminal domain. In yeast, this domain was found in another protein (Bdf1) that can interact with yTFIID (Matangkasombut et al, 2000). Most importantly, TAF4, instead of TAF1, was reported recently to play the most significant role in structural integrity of TFIID. This subunit was shown to form a core subcomplex of TFIID together with TAF5, TAF6, TAF9

and TAF12 *in vivo* (Wright et al, 2006). Therefore, in order to recombinantly produce hTAF1, this subunit would probably have to be co-expressed with the core TFIID subcomplex containing TAF4, TAF5, TAF6, TAF9 and TAF12.

Biacore experiments presented in this chapter showed that hTAF2 directly interacts with the hTAF8/10 complex. The dissociation constant of hTAF2-hTAF8/10 complex was tentatively calculated to be 101 nM, indicating a moderately strong interaction. To measure the dissociation constant reliably, it is necessary to obtain a better fit for the Biacore data. One possibility would be to calculate hTAF8/TAF10 dimer-tetramer equilibrium constant by sedimentation equilibrium ultracentrifugation analysis of the complex at different concentrations. Using this constant, the obtained Biacore data could be fitted into a different model, taking hTAF8/10 dimer and tetramer concentrations into account. This model would produce a dissociation constant for hTAF2-hTAF8/10 dimer and hTAF2-hTAF8/10 tetramer separately. Experiments using the hTAF2 core instead of the full-length protein indicated that this domain has hTAF8/10 binding activity, but dissociation constant could not be determined. An extensive screening of running buffer conditions, as performed for the full-length hTAF2, may provide better data that would allow the calculation of the hTAF2 core-hTAF8/10 complex dissociation constant. Nevertheless, together with the data on hTAF2 tail domain having no hTAF8/10 binding activity, this experiment points to the hTAF2 core as the hTAF8/10 interaction domain.

DNA band shift experiments with hTAF2 indicate that this protein has non-specific DNA binding activity with a dissociation constant in the micromolar range. Interestingly, the hTAF2-DNA complex was prone to aggregation and migrated into the native polyacrylamide gels only if supplemented with low SDS concentrations (0.05%) that generally do not cause protein denaturation (Gudiksen et al, 2006). The fact that non-ionic detergents could not prevent this aggregation indicates that aggregates were dependent both on hydrophobic and ionic interactions. In experiments where DNA was titrated with hTAF2, the bands were shifted further with higher protein concentrations. This behavior indicated that hTAF2 can form different stoichiometry complexes with DNA. These complexes were difficult to distinguish since the bands were smeared as expected for non-specific interaction where the protein-DNA complex is not distinct.

Experiments with hTAF2 domains have shown that DNA binding activity of hTAF2 is localized in the tail region. This domain contains a Lys-rich sequence that could explain the affinity for DNA. CD spectrum of the tail domain, which is mostly unstructured, changes

slightly in the presence of DNA, possibly indicating increased α -helix content. Only a fraction of protein molecules could be saturated with DNA. A more sensitive method such as NMR could give more insight into the hTAF2 tail structural changes as well point to the residues involved in DNA binding. The fact that the bands shifted by the tail domain were migrating more slowly than expected, comparable to the full-length protein, indicated that complex has greater than 1:1 stoichiometry. The high positive charge of the tail domain has also to be taken into account, since it might at least partially explain the slow mobility of the complex.

Taken together, the results of hTAF2 DNA binding experiments are in agreement with the model proposed for the hTFIID core promoter binding in which hTAF2 would only stabilize hTFIID binding on TATA-less promoters, but would not be responsible directly for the initiator element recognition (Martin et al, 1999). A protein that was proposed to carry out this latter function is hTAF1. To obtain direct evidence for this function, a purified TFIID core subcomplex with and without hTAF1 would be necessary, since hTAF1 is not stable on its own. If this complex could be produced recombinantly and bound with specific DNA, it would present a first step towards obtaining structural information at high resolution that could provide detailed insight into TFIID function in core promoter recognition and transcription in general.

6 REFERENCES

- Adachi, N. and Lieber, M.R. (2002) Bidirectional gene organization: A common architectural feature of the human genome. *Cell*, **109**, 807-809.
- Albright, S.R. and Tjian, R. (2000) TAFs revisited: More data reveal new twists and confirm old ideas. *Gene*, **242**, 1-13.
- Andel, F. 3rd, Ladurner, A. G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science*, **286**, 2153–2156.
- Aoyagi, N. and Wassarman, D.A. (2001) Developmental and transcriptional consequences of mutations in *Drosophila* TAF(II)60. *Mol Cell Biol*, **20**, 6808-6819.
- Arkhipova, I.R. (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics*, **139**, 1359-1369.
- Berger, I., Fitzgerald, D.J., Richmond, T.J. (2004) Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol*. **22**, 1583-1587.
- Berk, A.J. (2000) TBP-like factors come into focus. *Cell*, **103**, 5-8.
- Berroteran, R.W., Ware, D.E., Hampsey, M. (1994) The *sua8* suppressors of *Saccharomyces cerevisiae* encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to *sua7* (TFIIB) mutations. *Mol Cell Biol*, **14**, 226-237.
- BIATEchnology handbook, (1994), Biacore.
- Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209-213
- Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: A current view of enhancer action. *Science*, **281**, 60-63
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature*, **371**, 435-438
- Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999) Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. *Science*, **286**, 2151–2153.
- Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem*, **50**, 349-383.
- Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol*, **212**, 563-578.
- Bulger, M. and Groudine, M. (1999) Looping versus linking: Toward a model for long-distance gene activation. *Genes & Dev*, **13**, 2465-2477.

- Burke, T.W. and Kadonaga, J.T. (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev*, **11**, 3020-3031.
- Burke, T.W. and Kadonaga, J.T. (1996) *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev*, **10**, 711-724.
- Burley, S.K. and Roeder, R.G. (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem*, **65**, 769-799.
- Butler, J.E. and Kadonaga JT. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev*, **16**, 2583-2592.
- Carcamo, J., Buckbinder, L., and Reinberg, D. (1991) The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc. Natl. Acad. Sci*, **88**, 8052-8056.
- Cavallini, B., Faus, I., Matthes, H., Chipoulet, J. M., Winsor, B., Egly, J. M., and Chambon, P. (1989). Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. *Proc. Natl. Acad. Sci. USA*, **86**, 9803-9807.
- Chalkley, G.E., Verrijzer, C.P. (1999) DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J*. **18**, 4835-4845.
- Chen, J.L., Attardi, L.D., Verrijzer, C.P., Yokomori, K., Tjian, R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93-105.
- Chen, J.L., Tjian, R. (1996) Reconstitution of TATA-binding protein-associated factor/TATA-binding protein complexes for in vitro transcription. *Methods Enzymol*, **273**, 208-217.
- Cheriyath, V., Novina, C.D., and Roy, A.L. (1998) TFII-I regulates V β promoter activity through an initiator element. *Mol. Cell. Biol*, **18**, 4444-4454.
- Chi, T. and Carey, M. (1996) Assembly of the isomerized TFIIA--TFIID--TATA ternary complex is necessary and sufficient for gene activation. *Genes Dev*, **10**, 2540-2550.
- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) Promoter sequences of eukaryotic protein-coding genes. *Science*, **209**, 1406-1414.
- Cramer, P., Bushnell, D.A., Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863-76.
- Cramer P. (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol*, **12**, 89-97.
- Crick F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561-563.
- Crowley, T.E., Hoey, T., Liu, J.-K., Jan, Y.N., Jan, L.Y., and Tjian, R. (1993) A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature*, **361**, 557-561.
- Dantanel, J.-C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000) TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol Cell*, **6**, 715-722.
- Dynlacht, B. D., Hoey, T., and Tjian, R. (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell*, **66**, 563-576.

- Emami, K.H., Jain, A., Smale, S.T. (1997) Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. *Genes Dev*, **11**, 3007-19.
- Emami, K.H., Navarre, W.W. and Smale, S.T. (1995) Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. *Mol. Cell. Biol*, **15**, 5906-5916.
- Evans, R., Fairley, J.A. and Roberts, S.G.E. (2001) Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev*, **15**, 2945-2949.
- Fairley, J.A., Evans, R., Hawkes, N.A. and Roberts, S.G. (2002) Core promoter-dependent TFIIB conformation and a role for TFIIB conformation in transcription start site selection. *Mol Cell Biol*, **22**, 6697-6705.
- Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res*, **9**, 6505-6525.
- Gangloff, Y. G., Romier, C., Thuault, S., Werten, S. and Davidson, I. (2001) The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci*, **26**, 250-257.
- Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res*, **9**, 3047-3060.
- Geiduschek, E.P. and Bartlett, M. S. (2000) Engines of gene expression. *Nature Struct. Biol*, **7**, 437-439.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D. (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, **292**, 1876-1882.
- Goldberg, M.L. (1979) "*Sequence analysis of Drosophila histone genes.*" Ph.D. dissertation Stanford University, California.
- Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, JR 3rd, Workman, J.L. (1998) A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell*, **94**, 45-53.
- Grosschedl, R. and Birnstiel, M.L. (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. *Proc Natl Acad Sci U S A*, **77**, 1432-1436.
- Grueneberg, D.A., Henry, R.W., Brauer, A., Novina, C.D., Cheriya, V., Roy, A.L. and Gilman, M. (1997) A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes Dev*, **11**, 2482-2493.
- Gudiksen, K.L., Gitlin, I. and Whitesides, G.M. (2006) Differentiation of proteins based on characteristic patterns of association and denaturation in solutions of SDS. *Proc Natl Acad Sci U S A*, **103**, 7968-7972.
- Guermeur, Y. (PhD Thesis) *Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines.*

- Hansen, S.K., Takada, S., Jacobson, R.H., Lis, J.T. and Tjian, R. (1997) Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell*, **91**, 71-83.
- Hernandez N. (1993) TBP, a universal eukaryotic transcription factor? *Genes Dev*, **7**, 1291-1308.
- Holmes, M.C. and Tjian, R. (2000) Promoter-selective properties of the TBP-related factor TRF1. *Science*, **288**, 867-870.
- Horikoshi, M., Carey, M.F., Kakidani, H. and Roeder, R.G. (1988) Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell*, **54**, 665-669.
- Horikoshi, M., Hai, T., Lin, Y.S., Green, M.R. and Roeder, R.G. (1988) Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell*, **54**, 1033-1042.
- Hu, S.L. and Manley, J.L. (1981) DNA sequence required for initiation of transcription *in vitro* from the major late promoter of adenovirus 2. *Proc Natl Acad Sci U S A*, **78**, 820-824.
- Hultmark, D., Klemenz, R. and Gehring, W.J. (1986) Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*. *Cell*, **44**, 429-438.
- Indra, A.K., Mohan, W.S. 2nd, Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., Tora, L. (2005) TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. *Dev Biol*. **285**, 28-37.
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol*, **14**, 116-127.
- Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000) The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol Cell*, **6**, 705-713.
- Karlsson, R. (2004) SPR for molecular interaction analysis: a review of emerging application areas. *J Mol Recognit*, **17**, 151-161.
- Kaufmann, J. and Smale, S.T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes & Dev*, **8**, 821-829.
- Kaufmann, J., Ahrens, K., Koop, R., Smale, S.T., Muller, R. (1998) CIF150, a human cofactor for transcription factor IID-dependent initiator function. *Mol Cell Biol*. **18**, 233-239.
- Kaufmann, J., Verrijzer, C.P., Shao, J., Smale, S.T. (1996) CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev*. **10**, 873-886.
- Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520-527.
- Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512-520.
- Klug A. (2001) Structural biology. A marvellous machine for making messages. *Science*, **292**, 1844-1846.

- Kutach, A.K. and Kadonaga, J.T. (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.*, **20**, 4754-4764.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. and Ebright, R.H. (1998) New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes & Dev.*, **12**, 34-44.
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, **23**, 4051-60.
- Lee, T.I. and Young, R.A. (2000) Transcription of eukaryotic protein-coding genes. *Annu. Rev. Gene.*, **34**, 77-137.
- Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L. and Schultz P. (2002) Mapping histone fold TAFs within yeast TFIID. *EMBO J*, **21**, 3424-3433.
- Leurent, C., Sanders, S.L., Demeny, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., Schultz, P. (2004) Mapping key functional sites within yeast TFIID. *EMBO J*. **23**, 719-727.
- Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11**, Issue 11.
- Lo, K. and Smale, S.T. (1996) Generality of a functional initiator consensus sequence. *Gene*, **182**, 13-22.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
- Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994) Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes & Dev*, **8**, 2282-2292.
- Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Bryant, S.H. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res*, **33**, D192-196.
- Martin, J., Halenbeck, R., Kaufmann, J. (1999) Human transcription factor hTAF(II)150 (CIF150) is involved in transcriptional regulation of cell cycle progression. *Mol Cell Biol*, **19**, 5548-5556.
- Martinez E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol*, **50**, 925-947.
- Martinez, E., Chiang, C.-M., Ge, H., and Roeder, R.G. (1994) TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J*, **13**, 3115-3126.
- Martinez, E., Ge, H., Tao, Y., Yuan, C.X., Palhan, V., Roeder, R.G. (1998) Novel cofactors and TFIIA mediate functional core promoter selectivity by the human TAFII150-containing TFIID complex. *Mol Cell Biol*, **18**, 6571-6583.

- Matangkasombut, O., Auty, R. and Buratowski, S. (2004) Structure and function of the TFIID complex. *Adv Protein Chem*, **67**, 67-92.
- Matangkasombut, O., Buratowski, R.M., Swilling, N.W. and Buratowski, S. (2000) Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev*, **14**, 951-962.
- Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem*, **255**, 11992-11996.
- Metzger, D., Scheer, E., Soldatov, A., Tora L. (1999) Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes. *EMBO J*, **18**, 4823-4834.
- Mohan, W.S. Jr., Scheer, E., Wendling, O., Metzger, D., Tora, L. (2003) TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. *Mol Cell Biol*, **23**, 4307-4318.
- Muller, F. and Tora, L. (2004) The multicoloured world of promoter recognition complexes. *EMBO J*, **23**, 2-8.
- Nakajima, N., Horikoshi, M. and Roeder, R.G. (1988) Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol Cell Biol*, **8**, 4028-4040.
- Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G. and Burley, S.K. (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, **377**, 119-28.
- Oelgeschläger, T., Chiang, C.-M., and Roeder, R.G. (1996) Topology and reorganization of a human TFIID-promoter complex. *Nature*, **382**, 735-738.
- Ogryzko, V.V., Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.J., Howard, B.H., Qin, J. and Nakatani, Y. (1998) Histone-like TAFs within the PCAF histone acetylase complex. *Cell*, **94**, 35-44.
- Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev*, **10**, 2657-2683.
- O'Shea-Greenfield, A. and Smale, S.T. (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J Biol Chem*, **267**, 1391-402.
- Purnell, B.A., Emanuel, P.A. and Gilmour, D.S. (1994) TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev*, **8**, 830-842.
- Roy, A.L., Du, H., Gregor, P.D., Novina, C.D., Martinez, E. and Roeder, R.G. (1997) Cloning of an Inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. *EMBO J*, **16**, 7091-7104.
- Roy, A.L., Meisterernst, M., Pognonec, P. and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature*, **354**, 245-248.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning – A Laboratory Manual*.
- Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-5467.
- Sawadogo, M. and Roeder, R.G. (1985) Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell*, **43**, 165-75.
- Singer, V.L., Wobbe, C.R. and Struhl, K. (1990) A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev*, **4**, 636-645.
- Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. *Cell*, **57**, 103-113.
- Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*, **72**, 449-479.
- Soldatov, A., Nabirochkina, E., Georgieva, S., Belenkaja, T. and Georgiev, P. (1999) TAFII40 protein is encoded by the e(y)1 gene: biological consequences of mutations. *Mol Cell Biol*, **19**, 3769-3778.
- Soutoglou, E., Demeny, M.A., Scheer, E., Fienga, G., Sassone-Corsi, P., Tora, L. (2005) The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners. *Mol Cell Biol*, **25**, 4092-4104.
- Struhl, K. (1989) Molecular mechanisms of transcriptional regulation in yeast. *Annu Rev Biochem*, **58**, 1051-77.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y. et al. (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res*, **11**, 677-684.
- Sweetser, D., Nonet, M., Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A*, **84**, 1192-6.
- Takada, S., Lis, J.T., Zhou, S., and Tjian, R. (2000) A TRF1: BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell*, **101**, 459-469.
- Tanese, N., Pugh, B.F. and Tjian, R. (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev*, **5**, 2212-2224.
- Tora, L. (2002) A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev*, **16**, 673-675.
- Tsai, F.T.F. and Sigler, P.B. (2000) Structural basis of preinitiation complex assembly on human Pol II promoters. *EMBO J*, **19**, 25-36.
- Usheva, A. and Shenk, T. (1994) TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell*, **76**, 1115-1121.
- Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., McCawley, P. (1977) The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro*, **13**, 213-217.

- Verrijzer, C.P. (2001) Transcription factor IID-Not so basal after all. *Science*, **293**, 2010-2011.
- Verrijzer, C.P., Chen, J.L., Yokomori, K., Tjian R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115-1125.
- Verrijzer, C.P., Yokomori, K., Chen, J.L., Tjian R. (1994) Drosophila TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science*, **264**, 933-941.
- Voet, D., Voet, J.G. (1995). *Biochemistry*. John Wiley & Sons, Inc.
- Wassarman, D.A., Sauer, F. (2001) TAF(II)250: a transcription toolbox. *J Cell Sci*, **11**, 2895-2902.
- Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R. and Chambon, P. (1980) Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. *Proc Natl Acad Sci U S A*, **77**, 7024-7028.
- Weis, L. and Reinberg, D. (1997) Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol*, **17**, 2973-2984.
- West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: Many functions, many mechanisms. *Genes Dev*, **16**, 271-288.
- Wieczorek, E., Brand, M., Jacq, X. and Tora, L. (1998) Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. *Nature*, **393**, 187-191.
- Woychik, N.A. and Hampsey, M. (2002) The RNA polymerase II machinery: Structure illuminates function. *Cell*, **108**, 453-463.
- Wright, K.J., Marr, M.T. 2nd., Tjian, R. (2006) TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A*, **103**, 12347-12352.
- Xu, L.C., Thali, M. and Schaffner, W. (1991) Upstream box/TATA box order is the major determinant of the direction of transcription. *Nucleic Acids Res*, **19**, 6699-6704.
- Yatherajam, G., Zhang, L., Kraemer, S.M. and Stargell, L.A. (2003) Protein-protein interaction map for yeast TFIID. *Nucleic Acids Res*, **31**, 1252-1260.
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, **98**, 811-824.

APPENDIX I: DNA AND PROTEIN SEQUENCES

Human TAF2 amino acid and DNA sequence

The full-length gene (as expressed from pDiFBhTAF2) contains 3588 bp or 1196 amino acids (Mw=136539 Da).

```

1      ATGGACTACAAAGACGATGACGATAAAAAACAGGAAGAAAGGAGACAAGGGCTTTGAAAGC
1      M D Y K D D D D K N R K K G D K G F E S

61     CCAAGGCCATATAAATTAACCCATCAGGTCGTCTGCATCAACAACATAAATTTCCAGAGA
21     P R P Y K L T H Q V V C I N N I N F Q R

121    AAATCTGTTGTGGGATTTGTGGAAGTACTATATTTCCACAGTTGCAAACCTTGAATAGA
41     K S V V G F V E L T I F P T V A N L N R

181    ATCAAGTTGAACAGCAAACAGTGTAGAATATAACCGAGTAAGGATCAATGATTTAGAGGCT
61     I K L N S K Q C R I Y R V R I N D L E A

241    GCTTTTATTTATAATGACCCAACCTTGGAAGTTTGTCCACAGTGAATCAAACAGAGAAAC
81     A F I Y N D P T L E V C H S E S K Q R N

301    CTCAATTATTTTTCCAATGCTTATGCAGCTGCAGTTAGTGCTGTGGACCCTGATGCAGGA
101    L N Y F S N A Y A A A V S A V D P D A G

361    AATGGAGAACTTTGCATTAAGGTTCCATCAGAGCTATGGAAACACGTTGATGAGTTAAAG
121    N G E L C I K V P S E L W K H V D E L K

421    GTCCTGAAGATACACATCAATTTTTCTTTGGATCAGCCCAAAGGAGGTCTTCATTTTTGTG
141    V L K I H I N F S L D Q P K G G L H F V

481    GTACCCAGTGTAGAGGGAAGTATGGCAGAGAGAGGTGCTCATGTTTTCTCTTGTGGGTAT
161    V P S V E G S M A E R G A H V F S C G Y

541    CAAAATTCTACAAGATTTTTGGTTCCTTGTGTTGATTCACTCTGAATTGTGTACATGG
181    Q N S T R F W F P C V D S Y S E L C T W

601    AAATTAGAATTTACAGTAGATGCTGCAATGGTTGCTGTTTCTAATGGCGATTTGGTGGAG
201    K L E F T V D A A M V A V S N G D L V E

661    ACAGTGTATACTCATGATATGAGGAAGAAAACCTTTCCATTATATGCTTACCATTCCCTACA
221    T V Y T H D M R K K T F H Y M L T I P T

721    GCAGCGTCAAATATCTCCTTGGCCATTGGACCATTTGAAATACTGGTAGATCCATACATG
241    A A S N I S L A I G P F E I L V D P Y M

781    CATGAGGTTACTCATTTTTGTTTGCCTTCTTCCATTGCTGAAACATAACCACATCA
261    H E V T H F C L P Q L L P L L K H T T S

841    TACCTTCATGAAGTCTTTGAATTTTTATGAAGAAATTCTTACATGTCGTTACCCATACTCC
281    Y L H E V F E F Y E E I L T C R Y P Y S

```

901 TGTTTTAAGACTGTCTTCATTGATGAGGCTTATGTTGAAGTGGCTGCTTATGCTTCCATG
301 C F K T V F I D E A Y V E V A A Y A S M

961 AGCATTTTTAGCACAAATCTTTTACACAGTGCCATGATTATAGATGAGACACCTTTGACT
321 S I F S T N L L H S A M I I D E T P L T

1021 AGAAGGTGTTTAGCCCAATCCTTGGCCCAGCAGTTTTTTGGTTGTTTCATATCTAGAATG
341 R R C L A Q S L A Q Q F F G C F I S R M

1081 TCTTGGTCTGATGAATGGGTGCTGAAGGGAATTCAGGCTATATCTATGGACTTTGGATG
361 S W S D E W V L K G I S G Y I Y G L W M

1141 AAAAAAATTTTTGGTGTAAATGAGTACCGCCATTGGATTAAAGAGGAGCTAGACAAAATA
381 K K T F G V N E Y R H W I K E E L D K I

1201 GTGGCATATGAACTAAAAACTGGTGGGGTTTTACTACATCCCATATTTGGTGGAGGAAAA
401 V A Y E L K T G G V L L H P I F G G G K

1261 GAGAAGGATAATCCGGCTTCCCATCTACACTTTTCAATAAAGCATCCACATACACTGTCC
421 E K D N P A S H L H F S I K H P H T L S

1321 TGGGAATACTACACTATGTTTCAGTGTAAGCCACCTTGTGATGAGATTGATTGAAAAT
441 W E Y Y T M F Q C K A H L V M R L I E N

1381 AGGATCAGTATGGAATTTATGCTACAAGTTTTCAATAAACTGCTAAGTCTGGCTAGTACT
461 R I S M E F M L Q V F N K L L S L A S T

1441 GCTTCATCTCAGAAGTCCAGTCACATATGTGGAGTCAGATGTTGGTTTTCCACATCTGGG
481 A S S Q K F Q S H M W S Q M L V S T S G

1501 TTTTTAAATCCATTTCAAATGTCTCTGGCAAAGATATTCAGCCGTTAATAAAGCAGTGG
501 F L K S I S N V S G K D I Q P L I K Q W

1561 GTAGATCAGAGTGGAGTGGTAAAATTTTATGGAAGTTTTGCATTTAATAGAAAACGAAAT
521 V D Q S G V V K F Y G S F A F N R K R N

1621 GTCTTGGAACTGGAATAAAACAGGACTATACATCTCCTGGAACTCAAAAATACGTGGGA
541 V L E L E I K Q D Y T S P G T Q K Y V G

1681 CCACTTAAAGTGACAGTGCAGGAGTTAGATGGATCCTTCAATCATACTGCAAATTGAA
561 P L K V T V Q E L D G S F N H T L Q I E

1741 GAAAACAGCCTTAAACATGATATAACCTGCCATTCCAAAAGTAGAAGGAATAAAAAGAAA
581 E N S L K H D I P C H S K S R R N K K K

1801 AAAATCCCACTGATGAATGGAGAAGAAGTTGATATGGATCTTTCTGCAATGGATGCTGAT
601 K I P L M N G E E V D M D L S A M D A D

1861 TCCCCTTTGCTGTGGATAAGGATAGACCCAGATATGTCAGTATTGAGGAAGGTAGAATTT
621 S P L L W I R I D P D M S V L R K V E F

1921 GAGCAAGCTGATTTTATGTGGCAGTATCAGCTCCGCTATGAGAGAGATGTTGTTGCACAG
641 E Q A D F M W Q Y Q L R Y E R D V V A Q

1981 CAGGAATCCATTTTGGCTTTGGAAAAATCCCTACTCCAGCATCTCGGCTTGCCTCACT
661 Q E S I L A L E K F P T P A S R L A L T

2041 GATATATTAGAACAAGAGCAGTGTCTTCTACAGAGTAAGAATGTCAGCTTGCTTCTGTCTT
681 D I L E Q E Q C F Y R V R M S A C F C L

2101 GCAAAGATTGCAAATTCATGGGTGAGCACATGGACAGGACCACCAGCCATGAAGTCACTC
701 A K I A N S M V S T W T G P P A M K S L

2161 TTCACTAGGATGTTTTGTTGTAAAAGTTGTCCAAACATTGTGAAAACAAACAACCTTTATG
721 F T R M F C C K S C P N I V K T N N F M

2221 AGCTTTCAAAGCTATTTTCTACAGAAGACTATGCCAGTTGCAATGGCTTTATTAAGAGAT
741 S F Q S Y F L Q K T M P V A M A L L R D

2281 GTTCATAATCTTTGTCCTAAAGAAGTCTTAACGTTTATTTTAGACTTAATCAAGTACAAT
761 V H N L C P K E V L T F I L D L I K Y N

2341 GACAACGGGAAAAATAAGTTTTTCAGATAACTATTATCGTGCAGAAATGATTGATGCCCTG
781 D N G K N K F S D N Y Y R A E M I D A L

2401 GCCAACTCTGTTACACCTGCAGTCAGTGTGAATAATGAAGTTAGAACCTTTGGATAACTTA
801 A N S V T P A V S V N N E V R T L D N L

2461 AATCCTGATGTGCGACTCATTCTTGAAGAAATCACCAGATTTTTGAATATGGAAAAACTT
821 N P D V R L I L E E I T R F L N M E K L

2521 CTTCCGAGTTACAGGCATACCATCACTGTCTGAGTTGTTGAGAGCCATACGGGTACTTCAG
841 L P S Y R H T I T V S C L R A I R V L Q

2581 AAGAACGGACATGTGCCAAGTGATCCAGCTCTTTTTAAATCTTATGCTGAATATGGCCAC
861 K N G H V P S D P A L F K S Y A E Y G H

2641 TTTGTGGACATTAGGATAGCAGCTTTGGAAGCAGTTGTTGATTATACTAAAGTGGACAGA
881 F V D I R I A A L E A V V D Y T K V D R

2701 AGTTATGAAGAACTGCAATGGCTACTTAATATGATTTCAGAATGACCCTGTACCCTATGTA
901 S Y E E L Q W L L N M I Q N D P V P Y V

2761 AGGCATAAGATTCTCAACATGTTGACTAAGAACCCACCATTACTAAGAACATGGAGTCT
921 R H K I L N M L T K N P P F T K N M E S

2821 CCCTTATGCAATGAAGCCCTGGTAGATCAACTTTGGAACTTATGAATTCTGGTACTTCA
941 P L C N E A L V D Q L W K L M N S G T S

2881 CATGACTGGAGGTTACGGTGTGGTGTCTGTGGACTTGTACTTCACACTTTTTGGCCTCAGT
961 H D W R L R C G A V D L Y F T L F G L S

2941 AGACCTTCCTGTTTACCCTTGCCAGAGCTTGGGTTGGTTCTTAATCTAAAGGAGAAAAAA
981 R P S C L P L P E L G L V L N L K E K K

3001 GCTGTCTTGAATCCTACCATAATCCAGAGTCAGTAGCAGGCAACCAAGAAGCTGCAAAT
1001 A V L N P T I I P E S V A G N Q E A A N

3061 AATCCAAGCAGTCACCCACAGCTAGTTGGATTTTCAGAACCCTTTTTCCAGTTCTCAAGAT
1021 N P S S H P Q L V G F Q N P F S S S Q D

3121 GAGGAGGAGATTGATATGGATACTGTTTCATGATAGCCAGGCCTTCATTTCCCATCATTTA
1041 E E E I D M D T V H D S Q A F I S H H L

3181 AACATGCTTGAAGGCCGTCAACTCCAGGGCTCTCGAAATATCGGCCAGCTAGCTCCCGA
1061 N M L E R P S T P G L S K Y R P A S S R

3241 TCTGCTTTAATACCCAGCACTCAGCAGGCTGTGACAGCACACCCACCACAAAACCCAG
1081 S A L I P Q H S A G C D S T P T T K P Q

3301 TGGAGTTTGGAACTTGCACGGAAGGGAACAGGTAAAGAACAAGCACCTTTGGAGATGAGT

1101 W S L E L A R K G T G K E Q A P L E M S
3361 ATGCATCCAGCGGCAAGCGCTCCACTCTCAGTCTTTACTAAGGAATCTACAGCCTCCAAA
1121 M H P A A S A P L S V F T K E S T A S K
3421 CACAGTGACCACCATCACCACCATCACCATGAGCACAAAGAAAAAGAAGAAGCATAAA
1141 H S D H H H H H H H E H K K K K K K H K
3481 CATAAGCACAAACACAAGCATAAGCATGACAGTAAAGAAAAGGACAAGGAGCCTTTCACT
1161 H K H K H K H K H D S K E K D K E P F T
3541 TTCTCCAGCCCTGCCAGTGGCAGGTCTATTCGTTCTCCTTCCCTTTCAGACTGA
1181 F S S P A S G R S I R S P S L S D *

Human TAF1 amino acid and DNA sequence

The full-length gene (as expressed from pFBDOpolhTAF1) contains 5793 bp or 1931 amino acids (Mw= 219099 Da).

```

1      ATGGGTAACCATGACAAGCGACGATGGAAAAAGAATTTTCATAGCCGTCTCAGCAGCCAAC
1      M G N H D K R R W K K N F I A V S A A N

61     CGCTTTAAGAAAATCTCATCTCCGGGGCAGCTAGCTGGAGCCACCCGCAGTTCGAAAAA
21     R F K K I S S S G A A S W S H P Q F E K

121    GGCGCCGACGACGACGACGACAAGGGCTCCCATATGGGACCCGGCTGCGATTTGCTGCTG
41     G A D D D D D K G S H M G P G C D L L L

181    CGGACAGCAGCTACCATCACTGCTGCCGCCATCATGTCAGACACGGACAGCGACGAAGAT
61     R T A A T I T A A A I M S D T D S D E D

241    TCCGCTGGAGGCGGCCCATTTTCTTTAGCGGGTTTCCTTTTCGGCAACATCAATGGAGCC
81     S A G G G P F S L A G F L F G N I N G A

301    GGGCAGCTGGAGGGGAAAGCGTCTTGGATGATGAATGTAAGAAGCACTTGGCAGGCTTG
101    G Q L E G E S V L D D E C K K H L A G L

361    GGGGCTTTGGGGCTGGGCAGCCTGATCACTGAACTCACGGCAAATGAAGAATTGACCGGG
121    G A L G L G S L I T E L T A N E E L T G

421    ACTGACGGTGCCTTGGTAAATGATGAAGGGTGGGTTAGGAGTACAGAAGATGCTGTGGAC
141    T D G A L V N D E G W V R S T E D A V D

481    TATTCAGACATCAATGAGGTGGCAGAAGATGAAAGCCGAAGATACCAGCAGACGATGGGG
161    Y S D I N E V A E D E S R R Y Q Q T M G

541    AGCTTGCAGCCCCTTTGCCACTCAGATTATGATGAAGATGACTATGATGCTGATTGTGAA
181    S L Q P L C H S D Y D E D D Y D A D C E

601    GACATTGATTGCAAGTTGATGCCTCCTCCACCTCCACCCCGGGACCAATGAAGAAGGAT
201    D I D C K L M P P P P P P P G P M K K D

661    AAGGACCAGGATTCTATTACTGGTGAGAAAGTGGACTTCAGTAGTTTCTCTGACTCAGAA
221    K D Q D S I T G E K V D F S S S S D S E

721    TCTGAGATGGGACCTCAGGAAGCAACACAGGCAGAATCTGAAGATGGAAAGCTGACCCTT
241    S E M G P Q E A T Q A E S E D G K L T L

781    CCATTGGCTGGGATTATGCAGCATGATGCCACCAAGCTGTTGCCAAGTGTACAGAACTT
261    P L A G I M Q H D A T K L L P S V T E L

841    TTTCCAGAATTTGACCTGGAAAGGTGTTACGTTTTCTACGTCTTTTTGGACCAGGGAAG
281    F P E F R P G K V L R F L R L F G P G K

901    AATGTCCCATCTGTTTGGCGGAGTGCTCGGAGAAAGAGGAAGAAGAAGCACCGTGAGCTG
301    N V P S V W R S A R R K R K K K H R E L

961    ATACAGGAAGAGCAGATCCAGGAGGTGGAGTGCTCAGTAGAATCAGAAGTCAGCCAGAAG
321    I Q E E Q I Q E V E C S V E S E V S Q K

1021  TCTTTGTGGAACTACGACTACGCTCCACCACCACCTCCAGAGCAGTGTCTCTCTGATGAT
341    S L W N Y D Y A P P P P P P E Q C L S D D

```

1081 GAAATCACGATGATGGCTCCTGTGGAGTCCAAATTTTCCCAATCAACTGGAGATATAGAT
361 E I T M M A P V E S K F S Q S T G D I D

1141 AAAGTGACAGATACCAAACCAAGAGTGGCTGAGTGGCGTTATGGGCCTGCCCGACTGTGG
381 K V T D T K P R V A E W R Y G P A R L W

1201 TATGATATGCTGGGTGTCCCTGAAGATGGCAGTGGGTTTACTATGGCTTCAAACCTGAGA
401 Y D M L G V P E D G S G F D Y G F K L R

1261 AAGACAGAACATGAACCTGTGATAAAATCTAGAATGATAGAGGAATTTAGGAACTTGAG
421 K T E H E P V I K S R M I E E F R K L E

1321 GAAAACAATGGCACTGATCTTCTGGCTGATGAAAACCTCCTGATGGTGACACAGCTGCAT
441 E N N G T D L L A D E N F L M V T Q L H

1381 TGGGAGGATGATATCATCTGGGATGGGGAGGATGTCAAACACAAAGGGACAAAACCTCAG
461 W E D D I I W D G E D V K H K G T K P Q

1441 CGTGCAAGCCTGGCAGGCTGGCTTCCTTCTAGCATGACTAGGAATGCGATGGCTTACAAT
481 R A S L A G W L P S S M T R N A M A Y N

1501 GTTCAGCAAGGTTTTGCAGCCACTCTTGATGATGACAAACCTTGGTACTCCATTTTTCCC
501 V Q Q G F A A T L D D D K P W Y S I F P

1561 ATTGACAATGAGGATCTGGTATATGGACGCTGGGAGGACAATATCATTTGGGATGCTCAG
521 I D N E D L V Y G R W E D N I I W D A Q

1621 GCCATGCCCCGGCTGTTGGAACCTCCTGTTTTGACACTTGATCCCAATGATGAGAACCTC
541 A M P R L L E P P V L T L D P N D E N L

1681 ATTTTGGAAATTCCTGATGAGAAGGAAGAGGCCACCTCTAACTCCCCCTCCAAGGAGAGT
561 I L E I P D E K E E A T S N S P S K E S

1741 AAGAAGGAATCATCTCTGAAGAAGAGTGAATTCTCTTAGGGAAAACAGGAGTCATCAAG
581 K K E S S L K K S R I L L G K T G V I K

1801 GAGGAACCACAGCAGAACATGTCTCAGCCAGAAGTGAAGATCCATGGAATCTCTCCAAT
601 E E P Q Q N M S Q P E V K D P W N L S N

1861 GATGAGTATTATTATCCCAAGCAACAGGGTCTTCGAGGCACCTTTGGAGGGAATATTATC
621 D E Y Y Y P K Q Q G L R G T F G G N I I

1921 CAGCATTCAATTCCTGCTGTGGAATTACGGCAGCCCTTCTTTCCCACCCACATGGGGCCC
641 Q H S I P A V E L R Q P F F P T H M G P

1981 ATCAAACCTCCGGCAGTTCATCGCCACCTCTGAAAAAGTACTCATTTGGTGCACTTTCT
661 I K L R Q F H R P P L K K Y S F G A L S

2041 CAGCCAGGTCCCCACTCAGTCCAACCTTTGCTAAAGCACATCAAAAAAAGGCCAAGATG
681 Q P G P H S V Q P L L K H I K K K A K M

2101 AGAGAACAAGAGAGGCAAGCTTCAGGTGGTGGAGAGATGTTTTTTATGCGCACACCTCAG
701 R E Q E R Q A S G G G E M F F M R T P Q

2161 GACCTCACAGGCAAAGATGGTGATCTTATTCTTGCAGAATATAGTGAGGAAAATGGACCC
721 D L T G K D G D L I L A E Y S E E N G P

2221 TTAATGATGCAGGTTGGCATGGCAACCAAGATAAAGA ACTATTATAAACGGAAACCTGGA
741 L M M Q V G M A T K I K N Y Y K R K P G

2281 AAAGATCCTGGAGCACCAGATTGTAATATGGGGAAACTGTTTACTGCCATACATCTCCT

761 K D P G A P D C K Y G E T V Y C H T S P
2341 TTCCTGGGTTCTCTCCATCCTGGCCAATTGCTGCAAGCATTGAGAACAACCTTTTTTCGT
781 F L G S L H P G Q L L Q A F E N N L F R
2401 GCTCCAATTTATCTTCATAAGATGCCAGAACTGATTTCTTGATCATTCCGACAAGACAG
801 A P I Y L H K M P E T D F L I I R T R Q
2461 GGTTACTATATTCCGGAATTAGTGGATATTTTTGTGGTTGGCCAGCAGTGTCCCTTGTTT
821 G Y Y I R E L V D I F V V G Q Q C P L F
2521 GAAGTTCCTGGGCCTAACTCCAAAAGGGCCAATACGCATATTCCGAGACTTTCTACAGGTT
841 E V P G P N S K R A N T H I R D F L Q V
2581 TTTATTTACCGCCTTTTCTGGAAAAGTAAAGATCGGCCACGGAGGATACGAATGGAAGAT
861 F I Y R L F W K S K D R P R R I R M E D
2641 ATAAAAAAGCCTTTTCTTCCCATTCCAGAAAGCAGCATCCGGAAGAGGCTAAAGCTCTGC
881 I K K A F P S H S E S S I R K R L K L C
2701 GCTGACTTCAAACGCACAGGGATGGACTCAAACCTGGTGGGTGCTTAAGTCTGATTTTTCGT
901 A D F K R T G M D S N W W V L K S D F R
2761 TTACCAACGGAAGAAGAGATCAGAGCTATGGTGTCCACCAGAGCAGTGTGTGCTTATTAT
921 L P T E E E I R A M V S P E Q C C A Y Y
2821 AGCATGATAGCTGCAGAGCAACGACTGAAGGATGCTGGCTATGGTGAGAAATCCTTTTTTT
941 S M I A A E Q R L K D A G Y G E K S F F
2881 GCTCCAGAAGAAGAAAATGAGGAAGATTTCCAGATGAAGATTGATGATGAAGTTTCGCACT
961 A P E E E N E E D F Q M K I D D E V R T
2941 GCCCCTTGGAACACCACAAGGGCCTTCATTGCTGCCATGAAGGGCAAGTGTCTGCTAGAG
981 A P W N T T R A F I A A M K G K C L L E
3001 GTGACTGGGGTGGCAGATCCCACGGGGTGTGGTGAAGGATTCTCCTATGTGAAGATTCCA
1001 V T G V A D P T G C G E G F S Y V K I P
3061 AACAAACCAACACAGCAGAAGGATGATAAAGAACCGCAGCCAGTGAAGAAGACAGTGACA
1021 N K P T Q Q K D D K E P Q P V K K T V T
3121 GGAACAGATGCAGACCTTCGTGCGCTTTCCCTGAAAAATGCCAAGCAACTTCTACGTAAA
1041 G T D A D L R R L S L K N A K Q L L R K
3181 TTTGGTGTGCCTGAGGAAGAGATTA AAAAGTTGTCCCCTGGGAAGTGATTGATGTGGTG
1061 F G V P E E E I K K L S R W E V I D V V
3241 CGCACAATGTCAACAGAACAGGCTCGTTCTGGAGAGGGGCCATGAGTAAATTTGCCCGT
1081 R T M S T E Q A R S G E G P M S K F A R
3301 GGATCAAGGTTTTCTGTGGCTGAGCATCAAGAGCGTTACAAAGAGGAATGTCAGCGCATC
1101 G S R F S V A E H Q E R Y K E E C Q R I
3361 TTTGACCTACAGAACAAGGTTCTGTCATCAACTGAAGTCTTATCAACTGACACAGACAGC
1121 F D L Q N K V L S S T E V L S T D T D S
3421 AGCTCAGCTGAAGATAGTACTTTGAAGAAATGGGAAAGAACATTGAGAACATGTTGCAG
1141 S S A E D S D F E E M G K N I E N M L Q
3481 AACAGAAAACCAGCTCTCAGCTTTACGTGAACGGGAGGAACAGGAGCGGAAGGAACTA
1161 N K K T S S Q L S R E R E E Q E R K E L

3541 CAGCGAATGCTACTGGCAGCAGGCTCAGCAGCATCCGGAACAATCACAGAGATGATGAC
1181 Q R M L L A A G S A A S G N N H R D D D

3601 ACAGCTTCCGTGACTAGCCTTAACTCTTCTGCCACTGGACGCTGTCTCAAGATTTATCGC
1201 T A S V T S L N S S A T G R C L K I Y R

3661 ACGTTTCGAGATGAAGAGGGGAAAGAGTATGTTTCGCTGTGAGACAGTCCGAAAACCAGCT
1221 T F R D E E G K E Y V R C E T V R K P A

3721 GTCATTGATGCCTATGTGCGCATACGGACTACAAAAGATGAGGAATTCATTTCGAAAATTT
1241 V I D A Y V R I R T T K D E E F I R K F

3781 GCCCTTTTTGATGAACAACATCGGGAAGAGATGCGAAAAGAACGGCGGAGGATTCAAGAG
1261 A L F D E Q H R E E M R K E R R R I Q E

3841 CAACTGAGGCGGCTTAAGAGGAACCAGGAAAAGGAGAAGCTTAAGGGTCTCTCTGAGAAG
1281 Q L R R L K R N Q E K E K L K G P P E K

3901 AAGCCCAAGAAAATGAAGGAGCGTCTGACCTAAAAGTAAATGTGGGGCATGTGGTGCC
1301 K P K K M K E R P D L K L K C G A C G A

3961 ATGGACACATGAGGACTAACAATTTCTGCCCCCTCTATTATCAAACAATGCGCCACCT
1321 I G H M R T N K F C P L Y Y Q T N A P P

4021 TCCAACCCTGTTGCCATGACAGAAGAACAGGAGGAGGAGTTGGAAAAGACAGTCATTCAT
1341 S N P V A M T E E Q E E E L E K T V I H

4081 AATGATAATGAAGAACTTATCAAGGTTGAAGGGACCAAAATTTGTCTTGGGGAAACAGCTA
1361 N D N E E L I K V E G T K I V L G K Q L

4141 ATTGAGAGTGGGATGAGGTTTCGCAGAAAATCTCTGGTTCTCAAGTTTCCTAAACAGCAG
1381 I E S A D E V R R K S L V L K F P K Q Q

4201 CTTCTCCAAAGAAGAAACGGCGAGTTGGAACCACTGTTCACTGTGACTATTTGAATAGA
1401 L P P K K K R R V G T T V H C D Y L N R

4261 CCTCATAAGTCCATCCACCGGCGCCGCACAGACCCTATGGTGACGCTGTCTGTCATCTTG
1421 P H K S I H R R R T D P M V T L S S I L

4321 GAGTCTATCATCAATGACATGAGAGATCTTCCAAATACATACCCTTTCCACACTCCAGTC
1441 E S I I N D M R D L P N T Y P F H T P V

4381 AATGCAAAGGTTGTAAAGGACTACTACAAAATCATCACTCGGCCAATGGACCTACAAACA
1461 N A K V V K D Y Y K I I T R P M D L Q T

4441 CTCCGCGAAAACGTGCGTAAACGCCTCTACCCATCTCGGGAAGAGTTTCAGAGAGCATCTG
1481 L R E N V R K R L Y P S R E E F R E H L

4501 GAGCTAATTGTGAAAAATAGTGCAACCTACAATGGGCCAAAACACTCATTGACTCAGATC
1501 E L I V K N S A T Y N G P K H S L T Q I

4561 TCTCAATCCATGCTGGATCTCTGTGATGAAAAACTCAAAGAGAAAAGAAGACAAATTAGCT
1521 S Q S M L D L C D E K L K E K E D K L A

4621 CGCTTAGAGAAAGCTATCAACCCCTTGCTGGATGATGATGACCAAGTGGCGTTTTCTTTTC
1541 R L E K A I N P L L D D D D Q V A F S F

4681 ATTCTGGACAACATTGTCAACCAGAAAATGATGGCAGTTCCAGATTCTTGGCCATTTTCAT
1561 I L D N I V T Q K M M A V P D S W P F H

4741 CACCCAGTTAATAAGAAATTTGTTCCAGATTATTACAAAGTGATTGTCAATCCAATGGAT
1581 H P V N K K F V P D Y Y K V I V N P M D

4801 TTAGAGACCATAACGTAAGAACATCTCCAAGCACAAGTATCAGAGTCGGGAGAGCTTTCTG
1601 L E T I R K N I S K H K Y Q S R E S F L

4861 GATGATGTAAACCTTATTCTGGCCAACAGTGTTAAGTATAATGGACCTGAGAGTCAGTAT
1621 D D V N L I L A N S V K Y N G P E S Q Y

4921 ACTAAGACTGCCAGGAGATTGTGAACGTCTGTTACCAGACATTGACTGAGTATGATGAA
1641 T K T A Q E I V N V C Y Q T L T E Y D E

4981 CATTGACTCAACTTGAGAAGGATATTTGTACTGCTAAAGAAGCAGCTTTGGAGGAAGCA
1661 H L T Q L E K D I C T A K E A A L E E A

5041 GAATTAGAAAGCCTGGACCCAATGACCCAGGGCCCTACACGCCTCAGCCTCCTGATTTG
1681 E L E S L D P M T P G P Y T P Q P P D L

5101 TATGATACCAACACATCCCTCAGTATGTCTCGAGATGCCTCTGTATTTCAAGATGAGAGC
1701 Y D T N T S L S M S R D A S V F Q D E S

5161 AATATGTCTGTCTTGGATATTTCCAGTGCCACTCCAGAAAAGCAGGTAACACAGGAAGGT
1721 N M S V L D I P S A T P E K Q V T Q E G

5221 GAAGATGGAGATGGTGTCTTGCAGATGAAGAGGAAGGAAGTGTACAACAGCCTCAAGCC
1741 E D G D G D L A D E E E G T V Q Q P Q A

5281 AGTGTCTGTATGAGGATTTGCTTATGTCTGAAGGAGAAGATGATGAGGAAGATGCTGGG
1761 S V L Y E D L L M S E G E D D E E D A G

5341 AGTGATGAAGAAGGAGACAATCCTTTCTCTGCTATCCAGCTGAGTGAAAGTGAAGTGAC
1781 S D E E G D N P F S A I Q L S E S G S D

5401 TCTGATGTGGGATCTGGTGAATAAGACCCAAACAACCCCGCATGCTTCAGGAGAACACA
1801 S D V G S G G I R P K Q P R M L Q E N T

5461 AGGATGGACATGGAATGAAGAAAGCATGATGTCCTATGAGGGAGACGGTGGGGAGGCT
1821 R M D M E N E E S M M S Y E G D G G E A

5521 TCCCATGGTTTGGAGGATAGCAACATCAGTTATGGGAGCTATGAGGAGCCTGATCCCAAG
1841 S H G L E D S N I S Y G S Y E E P D P K

5581 TCGAACACCCAAGACACAAGCTTCAGCAGCATCGGTGGGTATGAGGTATCAGAGGAGGAA
1861 S N T Q D T S F S S I G G Y E V S E E E

5641 GAAGATGAGGAGGAGGAAGAGCAGCGCTCTGGGCCGAGCGTACTAAGCCAGGTCCACCTG
1881 E D E E E E E Q R S G P S V L S Q V H L

5701 TCAGAGGACGAGGAGGACAGTGAGGATTTCCACTCCATTGCTGGGGACAGTGACTTGGAC
1901 S E D E E D S E D F H S I A G D S D L D

5761 TCTGATGAACGGTCCGGCCACCATCATCACCACCATTGA
1921 S D E R S G H H H H H H *

Human TBP core amino acid and DNA sequence

The TBP core gene (as expressed from pET28a and pFBDOtriple) contains 543 bp or 181 amino acids (Mw=20401 Da).

```

1      ATGTCTGGGATTGTACCGCAGCTGCAAAATATTGTATCCACAGTGAATCTTGGTTGTAAA
1      M S G I V P Q L Q N I V S T V N L G C K

61     CTTGACCTAAAGACCATTGCACTTCGTGCCCGAAACGCCGAATATAATCCCAAGCGGTTT
21     L D L K T I A L R A R N A E Y N P K R F

121    GCTGCGGTAATCATGAGGATAAGAGAGCCACGAACCACGGCACTGATTTTCAGTTCTGGG
41     A A V I M R I R E P R T T A L I F S S G

181    AAAATGGTGTGCACAGGAGCCAAGAGTGAAGAACAGTCCAGACTGGCAGCAAGAAAATAT
61     K M V C T G A K S E E Q S R L A A R K Y

241    GCTAGAGTTGTACAGAAGTTGGGTTTTCCAGCTAAGTTCTTGGACTTCAAGATTCAGAAC
81     A R V V Q K L G F P A K F L D F K I Q N

301    ATGGTGGGGAGCTGTGATGTGAAGTTTCCTATAAGGTTAGAAGGCCTTGTGCTCACCCAC
101    M V G S C D V K F P I R L E G L V L T H

361    CAACAATTTAGTAGTTATGAGCCAGAGTTATTTCTGGTTTAATCTACAGAATGATCAAAA
121    Q Q F S S Y E P E L F P G L I Y R M I K

421    CCCAGAATTGTTCTCCTTATTTTTGTTTCTGGAAAAGTTGTATTAACAGGTGCTAAAGTC
141    P R I V L L I F V S G K V V L T G A K V

481    AGAGCAGAAATTTATGAAGCATTGAAAACATCTACCCTATTCTAAAGGGATTCAGGAAG
161    R A E I Y E A F E N I Y P I L K G F R K

541    ACGACGTAA
181    T T *

```

Human TAF8/10 amino acid and DNA sequence

The full-length TAF8 gene (with a C-terminal His-tag) contains 963 bp or 321 amino acids (Mw= 35499 Da).

```

1      ATGGGTAACCACGCCGACGCGGCGGCCACAGCTGGGGCCGGTGGCTCCGGAACGAGATCG
1      M G N H A D A A A T A G A G G S G T R S

61     GGAAGTAAACAGTCCACTAACCCCTGCCGATAACTATCATCTGGCCCGGAGGAGAACCCTG
21     G S K Q S T N P A D N Y H L A R R R T L

121    CAGGTGGTTGTGAGCTCCTTGCTGACAGAGGCAGGGTTTGAGAGTGCCGAGAAAGCATCC
41     Q V V V S S L L T E A G F E S A E K A S

181    GTGAAACGCTGACAGAGATGCTGCAGAGCTACATTTTCAGAAATTGGGAGAAGTGCCAAG
61     V E T L T E M L Q S Y I S E I G R S A K

241    TCTTACTGTGAGCACACAGCCAGGACCCAGCCCACACTGTCCGATATCGTGGTTCACACTT
81     S Y C E H T A R T Q P T L S D I V V T L

301    GTTGAGATGGGTTTCAATGTGGACACTCTCCCTGCTTATGCAAACGGTCTCAGAGGATG
101    V E M G F N V D T L P A Y A K R S Q R M

361    GTCATCACTGCTCCTCCCGTGACCAATCAGCCAGTGACCCCCAAGGCCCTCACTGCAGGG
121    V I T A P P V T N Q P V T P K A L T A G

421    CAGAACCGACCCCACCCGCGCACATCCCCAGCCATTTTCTGAGTTCCTGATCCCCAC
141    Q N R P H P P H I P S H F P E F P D P H

481    ACCTACATCAAACTCCGACGTACCGTGAGCCCGTGTGAGACTACCAGGTCCTGCGGGAG
161    T Y I K T P T Y R E P V S D Y Q V L R E

541    AAGGCTGCATCCCAGAGGCGCGATGTGGAGCGGGCACTTACCCGTTTCATGGCCAAGACA
181    K A A S Q R R D V E R A L T R F M A K T

601    GGCGAGACTCAGAGTCTTTTCAAAGATGACGTCAGCACATTTCTTTGATTGCTGCCAGA
201    G E T Q S L F K D D V S T F P L I A A R

661    CCTTTCACCATCCCCCTACCTGACAGCTCTTCTTCCGTCTGAACTGGAGATGCAACAAATG
221    P F T I P Y L T A L L P S E L E M Q Q M

721    GAAGAGACAGATTCCCTCGGAGCAGGATGAACAGACAGACACAGAGAACCTTGCTCTTCAT
241    E E T D S S E Q D E Q T D T E N L A L H

781    ATCAGCATGGAGGATTCTGGAGCCGAGAAGGAGAACACCTCTGTCCTGCAGCAGAACCCC
261    I S M E D S G A E K E N T S V L Q Q N P

841    TCCTTGTCGGGTAGCCGGAATGGGGAGGAGAACATCATCGATAACCCTTATCTGCGGCCG
281    S L S G S R N G E E N I I D N P Y L R P

901    GTGAAGAAGCCCAAGATCCGCAGGAAGAAGTCCCTCTCCCGGTCCGGCCACCATCATCAC
301    V K K P K I R R K K S L S R S G H H H H

961    CACCATTGA
321    H H *

```

The full-length TAF10 gene contains 651 bp or 217 amino acids (Mw= 21568 Da).

```
1      ATGAGCTGCAGCGGCTCCGGCGCGGACCCCGAGGCGGGCGCCGGCCTCCGCCGCCTCGGCC
1      M S C S G S G A D P E A A P A S A A S A

61     CCGGGCCCCGCGCCCCCGGTCTCGGCTCCCGCCGCGCTGCCCTCCAGCACCGCCGCGGAG
21     P G P A P P V S A P A A L P S S T A A E

121    AACAAAGGCCAGCCCCGCGGGGACAGCGGGGGACCTGGGGCTGGAGCAGCTGCTGGGGGC
41     N K A S P A G T A G G P G A G A A A G G

181    ACGGGACCCTTGGCGGCGCGGGCCGGGGAGCCAGCTGAGCGGCGTGGGGCGGCTCCGGTG
61     T G P L A A R A G E P A E R R G A A P V

241    TCGGCGGGTGGCGGCGCGCCCCGGAGGGGGCCATATCTAACGGGGTTTACGTACTGCCG
81     S A G G A A P P E G A I S N G V Y V L P

301    AGCGCGGCCAACGGAGACGTGAAGCCCGTGGTGTCCAGCACGCCTTTGGTGGACTTCTTG
101    S A A N G D V K P V V S S T P L V D F L

361    ATGCAGCTGGAAGATTACACGCCTACGATCCCAGATGCAGTGACTGGTTACTACCTGAAC
121    M Q L E D Y T P T I P D A V T G Y Y L N

421    CGTGCTGGCTTTGAGGCCTCAGACCCACGCATAATTCGGCTCATCTCCTTAGCTGCCCAG
141    R A G F E A S D P R I I R L I S L A A Q

481    AAATTCATCTCAGATATTGCCAATGATGCCCTACAGCACTGCAAAATGAAGGGCACGGCC
161    K F I S D I A N D A L Q H C K M K G T A

541    TCCGGCAGCTCCCGGAGCAAGAGCAAGGACCGCAAGTACACTCTAACCATGGAGGACTTG
181    S G S S R S K S K D R K Y T L T M E D L

601    ACCCCTGCCCTCAGCGAGTATGGCATCAATGTGAAGAAGCCGCACTACTTCACCTGA
201    T P A L S E Y G I N V K K P H Y F T *
```


APPENDIX II: EXPRESSION OF TAF2 CONSTRUCTS IN *E. COLI*

Prokaryotic expression systems have an advantage over the baculovirus system because they are much less time consuming, easier to handle and frequently give high yield of the recombinant protein. Therefore, for the purpose of obtaining a crystallizable piece of hTAF2, the attempt was made to produce the fragments of this protein in *E. coli*. The chosen pieces of hTAF2 were based on the following tryptic fragments observed in the limited proteolysis analysis (section 4.2.3): TR-2, TR-4 and TR-5. The TAF2 “tail” domain was also produced in *E. coli* (details of this work are described in the section 4.4.1). Bacterial expression of the full-length hTAF2 was never attempted since the protein was assumed to be too large to be a reasonably good candidate for this system.

TR-2 (TAF2 “core” domain), TR-4 and TR-5 fragments were cloned into a pET28a expression vector in 2 steps. First, the PCR product was obtained using the appropriate primers (Table A.1), containing the *Nco* I and *Xho* I restriction sites. pDiFBTAF2 plasmid was used as a template for the PCR reaction. The primers were designed to remove the FLAG tag and the enterokinase site from the protein’s N-terminus and to allow the vector encoded His-tag to be added to the C-terminus of each construct. The PCR products were cloned into the pCR2.1-TOPO vector, and, as a second step, subcloned into pET28a using the *Nco* I and *Xho* I restriction sites. The final constructs were verified by sequencing. The resulting protein sequences are listed in the Appendix I.

CONSTRUCT	SIZE	AMINO ACID POSITION (App. I)	CLONING PRIMERS	DESCRIPTION
pET28aTR-2	114.1 kDa	14 - 999	TAF2An TAF2AcBact	based on TR-2 (TAF2 core domain)
pET28aTR-4	67.7 kDa	14 - 596	TAF2An TAF2Tr4c	based on TR-4 (aminopeptidase domain)
pET28aTR-5	47.5 kDa	597 - 999	TAF2Tr5n TAF2AcBact	based on TR-5

TABLE A.1. TAF2 constructs for bacterial expression.

For the expression tests, *E. coli* Rosetta strain was transformed with pET28aTR-2, pET28aTR-4 or pET28aTR-5 construct. The cells were grown in a small scale liquid culture to OD₆₀₀ of 0.4-0.6 and induced with 0.8 mM IPTG. The induced cultures were incubated for 2.5 h at 37° C. The samples of induced and uninduced control culture were taken for the SDS-PAGE analysis. Solubility test was done by resuspending the cell pellets in Bacterial lysis buffer, followed by a 10 s sonication and centrifugation to separate the soluble from the insoluble fraction (as described in the section 2.3.3).

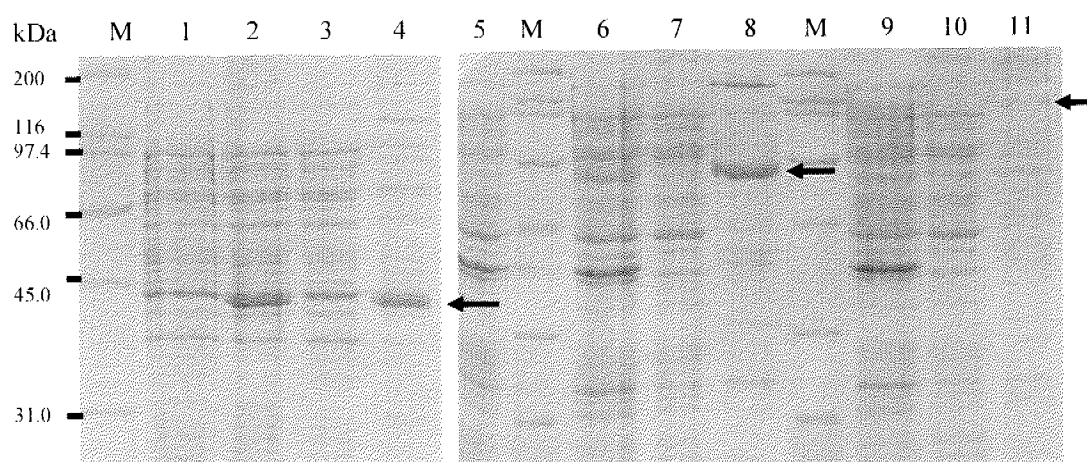


FIGURE A.1. Expression of hTAF2 constructs in *E. coli*.

Lanes 1 and 5 – uninduced sample. Lanes 2, 3, 4 – whole cell extract, soluble fraction, insoluble fraction from pET28aTR-5, respectively. Lanes 6, 7, 8 – whole cell extract, soluble fraction, insoluble fraction from pET28aTR-4, respectively. Lanes 9, 10, 11 – whole cell extract, soluble fraction, insoluble fraction from pET28aTR-2, respectively. The arrows point to the TAF2 constructs position. Lane “M” contains the prestained protein marker (BioRad).

Figure A.1 shows the results of the experiment: all constructs were successfully expressed and all of them were found in the insoluble fraction. The expression was also attempted at 18° C (overnight) and in the SPM media (as described for human TBPC expression in section 2.5.5). None of these strategies, which generally slow down protein production, were successful for obtaining soluble proteins, indicating that they were not properly folded. Therefore, the work with these constructs was discontinued. There could be several reasons for insolubility: inability of the *E. coli* cellular machinery to stop these proteins from aggregating during translation and folding (for example, lack of a specific chaperone), absence of a post-translational modification specific to eukaryotic cells necessary for proper folding, or the

structure of the constructs themselves. In case of TR-4 and TR-5, as indicated by the results of the experiments described in the section 4.3.2, these fragments might not be an autonomous folding unit (structural domain) within hTAF2. Therefore, they might require the presence of another protein piece to fold correctly.

APPENDIX III: TAF2 CRYSTALLIZATION ATTEMPTS – SUMMARY AND PERSPECTIVES

During this work, extensive crystallization experiments were conducted with human TAF2. Crystallization methods that were used are described in detail in the section 2.9.

Crystallization setups have been made with both the full-length version of hTAF2, purified as described in the Chapter 3, as well as the hTAF2 core (116 kDa construct A, described in the Chapter 4). Purified material has been concentrated generally between 5 and 10 mg/ml and mixed with the reservoir solution in ratios 1:1, 0.5:1 and 1:2. The Cartesian robot (operated by Beat Blattman, University of Zürich) was used for the majority of the crystal setups. All available commercial crystallization screens, listed in section 2.9.4, have been used as well as one customized screen. Setups were stored at both 4° C and 20° C. Unfortunately, none of these conditions produced protein crystals.

However, limited proteolysis experiments described in the Chapter 4 provide guidelines for future crystallization trials. Tryptic digests of the TAF2 core, purified over a Resource S cation exchange column, show that there is a peptide that can be separated from the rest of the core using a salt gradient. Since it can be easily separated, this peptide does not belong to the core hTAF2 structure and could interfere with the crystal formation. One direction for the future hTAF2 crystallization is setting up purified tryptic mixtures. Identification of the missing peptide would allow construct design for recombinant co-expression of tryptic fragments, providing crystallization material of higher purity and homogeneity.

CURRICULUM VITAE

Name: Martina Mijušković

Date and place of birth: March 26, 1979 in Bjelovar, Croatia

Citizenship: Croatian

Education:

2002-2007: Ph.D. studies at the Institute of Molecular Biology and Biophysics, ETH Zurich, Switzerland.

Ph.D. thesis: "Structural and Biochemical Studies on Human TAF2"

Supervisor: Prof. T. J. Richmond

2004-2006 Teaching assistant for 'Molecular Biology and Biophysics I', ETH Zurich

1997-2002: Undergraduate studies of Molecular Biology, Faculty of Science, University of Zagreb, Croatia.

Diploma thesis at the Faculty of Science, Department of Chemistry: "Unusual Seryl-tRNA Synthetases from Methanogenic Archaea"

Supervisor: Prof. Ivana Weygand-Durasevic

1993-1997: Grammar school, Bjelovar, Croatia.

Music high school, Bjelovar, Croatia.

1986-1993: Elementary music school, Bjelovar, Croatia.

1985-1993: General elementary school, Bjelovar, Croatia.