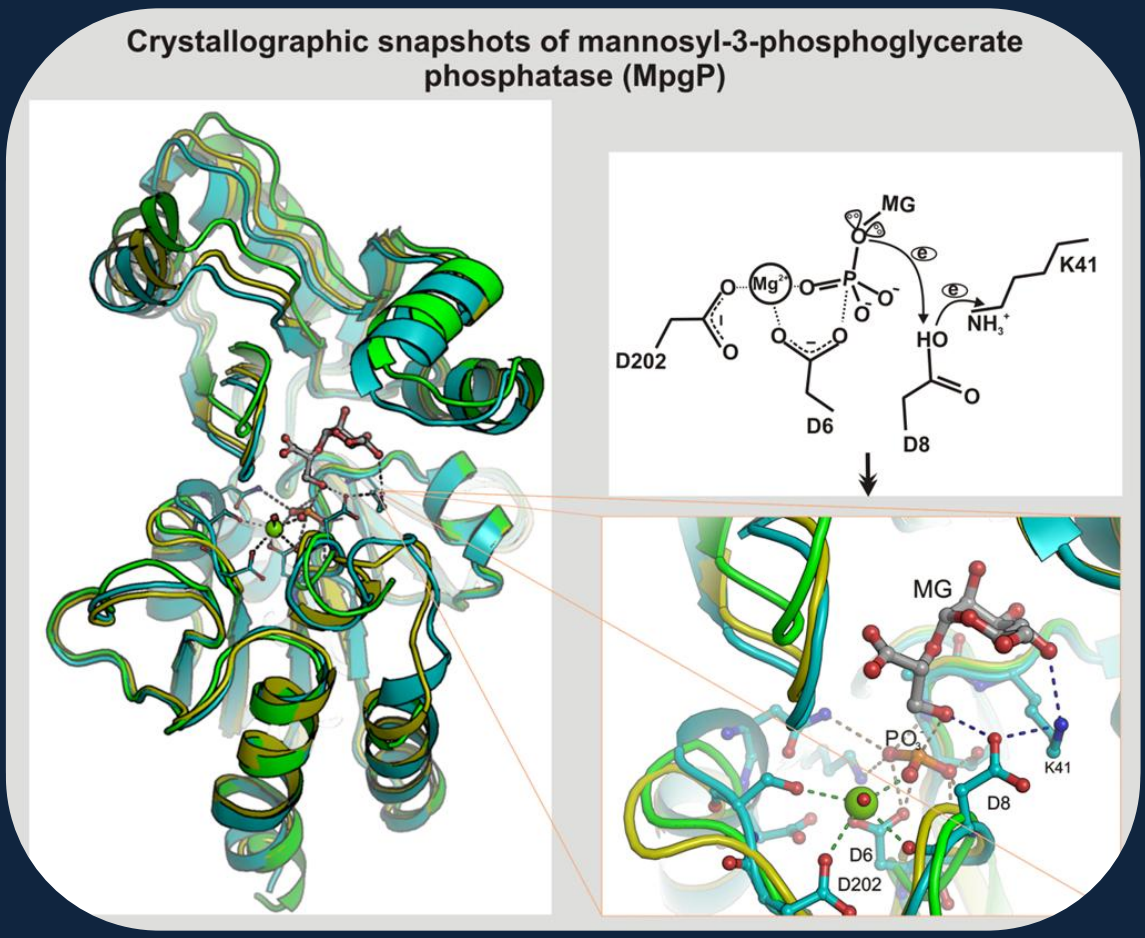
Structure-function relationships in a glycosyltransferase, a phosphatase and an oxidoreductase

Susana Margarida Pires Gonçalves



Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
May, 2012

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL

Knowledge Creation

Structure-function relationships in a glycosyltransferase, a phosphatase and an oxidoreductase

Susana Margarida Pires Gonçalves

Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, May, 2012

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

A thesis submitted in conformity with the requirements for the degree of Doctor of Instituto de Tecnologia Química Biológica, Universidade Nova de Lisboa.

Supervisor: Dr. Pedro Marques Matias

A António e Arcília, meus pais.

ACKOWLEDGMENT

I wish to express my sincere gratitude to my supervisor Dr. Pedro Matias for his careful guidance, his extended patience, his invariable availability and most of all his encouragement throughout my doctoral studies.

I extend my appreciation to my advisory committee; Dr. Helena Santos and Dr. Peter Donner, and to Dr. Maria Arménia Carrondo, the three of whom provided me with sound advice.

Special thanks to Dr. Antony Dean for its valuable recommendations on the elucidation of the catalytic mechanism of β-oxidative dehydrogenation by isocitrate dehydrogenase, as well as for his support and encouragement.

Dr. Nuno Borges and Dr. Helena Santos, in helpful discussions of the biochemistry behind the "two-step" pathway of mannosylglycerate.

I wish to extend my appreciation to my labmates for their friendship and for their precious hints on science and on life.

I am profoundly grateful to my parents, my brother and sisters, who played fundamental role in every aspects of my life, and particularly, in my persuit for accomplishing my PhD thesis.

THESIS ABSTRACT

Enzyme evolution is often constrained by aspects of catalysis. Mechanistically diverse enzymes evolved from a common ancestor still preserve those structural signatures essential to the core chemistry retained by all members of the superfamily. Indeed, these shared features allow superfamilies to be accurately classified, while derived features allow nested families and subfamilies to be identified in a hierarchical fashion. Accurate classification has helped elucidate mechanisms promoting functional diversification, for example catalytic promiscuity, and protein engineering by rational design.

Nowadays, a holistic view of enzymes` regulatory mechanisms and catalytic proficiency is provided by the identification of conserved features of molecular architecture in combination with aspects of reaction dynamics.

My work focused on the structural elucidation and analysis of three enzymes: a glycosyltransferase; a phosphatase and an oxidorreductase. “Snapshots” along the reaction coordinate of each enzyme were obtained by combining X-ray diffraction with “cryo-trapping” ligand-binding methods. These were used to characterize the molecular mechanisms involved in substrate recognition and binding. They were also used to distinguish between models proposed for the catalytic mechanisms of each enzyme, and provide insights into enzyme dynamics essential for catalysis and the stereo and regio-selective strategies at work.

The structural elucidation of the “two-step pathway” of mannosylglycerate (MG) in *Thermus thermophilus* HB27 was carried out as part of a broader study on the molecular mechanisms of adaptation to extreme environments. MG is a canonical sugar-derivative compatible solute that protects cells against osmotic and thermal stresses. In the first step, mannosyl-3-phosphoglycerate synthase (MpgS; E.C. 2.4.1.217) catalyses the transfer of the mannosyl-moiety from the activated precursor GDP-α-D-mannose to the activated nucleophile acceptor D-glycerate-3-phosphate, yielding mannosyl-3-

phosphoglycerate. In the second step, mannosyl-3-phosphoglycerate is dephosphorylated to MG by mannosyl-3-phosphoglycerate phosphatase (MpgP; E.C. 3.1.3.70). The third enzyme studied is the metal dependent $NADP^+$-linked isocitrate dehydrogenase (IDH; E.C.1.1.1.42) from *Escherichia coli*. IDH catalyses the oxidative decarboxylation of isocitrate to α-ketoglutarate. *E. coli* IDH is one of the pioneering case-studies of concerted protein dynamics with a productive Michaelis complex. However, a complete picture of these events has been hampered by limitations at the level of crystallographic packing.

Main conclusions from each study are:

a) Structural elucidation shows *T. thermophilus* HB27 MpgS is a GT55 glycosyltransferase, a family that falls within the retaining GT-A enzymes (www.cazy.org). Two flexible loops involved in key interactions for the productive binding of substrates were identified by comparing the apo-form with the metal-substrate complex [Mg^{2+}:GDP-α-D-mannose]. A second metal binding site was found about 6 Å away from the mannose moiety. Kinetic and mutagenesis studies provided evidence that this second metal site is indispensable for catalysis. Additionally, Asp167 of the conserved D-X-D motif was proposed as the catalytic nucleophile in light of the current mechanistic models for the retaining GTs. A survey of enzymes with the GT-A fold and with a bound sugar-donor, or a bound analogue, was used to identify the orientation of the scissile glycosidic bond of the sugar donor with respect to the D/E-X-D motif. The glycosidic bond is oriented towards the motif in the retaining GTs and away from it in the inverting GTs. This feature, which explains the stereochemistry of the reactions, provides a structural signature that will assist efforts in classifying the inverting and retaining GTs.

b) Structural elucidation shows *T. thermophilus* HB27 MpgP is a metal-dependent Haloalcanoic Acid Dehalogenase-like phosphatase (HAD-like) belonging to the Mannosyl-3-Phosphoglycerate Phosphatase family of the Cof-type phosphatases. This structure provides one of the best examples of the combinatorial nature of functional core units, allowing specialization towards

substrates while preserving core reaction chemistry. The domains of the “open” apo- and the “closed” holo- forms of MpgP are related to each other by a hinge rotation. The dynamics of the catalytic machinery were followed using cryo-trapped reaction species to provide “crystallographic snapshots” along the reaction cycle. Results suggest that phosphoryl-transfer by MpgP from *T. thermophilus* HB27 involves a concerted D_NA_N mechanism with Asp8 acting as a catalytic acid in the formation of a short-lived metaphosphate intermediate that is immediately subjected to nucleophilic attack by water. These structures identify the principle mechanistic features of phosphoryl monoester transfer catalysis in members of the HAD superfamily. More generally, they suggest a possible continuum of phosphoryl transfer mechanisms, ranging from those that are purely associative to those that are purely dissociative.

c) Comparison of the crystallographic structures of a *pseudo*-Michaelis-Menten complex of the wildtype *Eco*IDH, the K100M mutant trapped with its reaction products and several other *Eco*IDH structures identifies three distinct conformational states: open, *quasi*-closed and closed. Structural comparisons suggest substrate binding initiates domain closure, with hinge dynamics that span the central β-sheet of each monomer in the biological homodimer. Conserved catalytic residues binding the nicotinamide ring of the $NADP^+$ coenzyme and the metal-bound substrate move as rigid bodies together with the hinge rotation. The closed conformations of both the wildtype *pseudo*-Michaelis ternary complex and the K100M reaction product complex reveal, for the first time, a realistic picture of the mechanistic details of the oxidative β-decarboxylation reaction. The structures are entirely in accordance with the postulated roles for the Try160-Lys230* diad in Brønsted acid-base catalysis, as well as revealing the proton relay essential to catalysis.

SUMÁRIO DA TESE

A evolução enzimática é na maioria dos casos determinada por aspectos associados à catálise. Enzimas que são funcionalmente distintas, mas que partilham uma origem ancestral comum, preservam assinaturas estruturais que asseguram a viabilidade da química base, comum a todos os membros de uma super-família. Estas assinaturas são fundamentais para a classificação de enzimas em diferentes super-famílias, assim como variantes destas assinaturas são indicações para a hierarquização de membros homólogos em famílias e sub-famílias. Tal sistema permite por sua vez a identificação dos mecanismos responsáveis pela diversidade funcional, entre os quais os factores estruturais que favorecem a promiscuidade catalítica e que são relevantes para a indústria biotecnológica e na engenharia de proteínas. Hoje em dia, os conceitos de estrutura e dinâmica estão correlacionados com o de catálise enzimática através da identificação de um conjunto de “movimentos-chave” com uma arquitectura molecular típica.

O meu trabalho centrou-se na elucidação da estrutura tridimensional e na análise estrutural de três enzimas: uma glicosiltransferase, uma fosfatase e uma oxidorreductase. Combinando técnicas de difracção de raios-X com métodos de crio-conservação de ligandos com diferentes tempos de incubação nos cristais das enzimas estudadas, obtiveram-se “instantâneos fotográficos” das diferentes etapas das coordenadas de reacção para cada enzima em estudo. A análise destas permitiu a identificação de mecanismos moleculares intervenientes no reconhecimento e ligação do substrato, a análise dos diferentes mecanismos de reacção com base nos modelos correntes, a aquisição de novos conceitos da relação entre a catálise e a dinâmica enzimáticas, e a identificação de determinantes que contribuem para a regio- e estereo-selectividades das enzimas estudadas.

A elucidação estrutural da via dos “dois-passos” para a síntese do manosilglicerato (MG) em *Thermus thermophilus* HB27 foi um dos subtópicos abordado, inserido na temática do estudo dos mecanismos de adaptação a

ambientes extremos. O MG é um soluto compatível canónico, da classe dos derivados de açúcares, e que protege as células contra as tensões osmótica e térmica. A enzima manosil-3-fosfoglicerato sintetase (MpgS; E.C. 2.4.1.217), catalisa a transferência do grupo manosilo da GDP-α-D-manose para o aceitador nucleofílico 3-D-fosfoglicerato, produzindo o intermediário manosil-3-fosfoglicerato (MPG). No segundo passo da reacção, o MPG é desfosforilado a MG pela manosil-3-fosfoglicerato fosfatase (MpgP; E.C. 3.1.3.70). O segundo tópico abordou novos avanços no estudo da isocitrato desidrogenase dependente de $NADP^+$ (IDH; E.C. 1.1.1.42) de *Escherichia coli.* A IDH catalisa a oxidação descarboxilativa do isocitrato em α-cetoglutarato e é um dos casos de estudo pioneiros na demonstração da acção concertada entre dinâmica e catálise enzimáticas. No entanto, limitações devidas ao arranjo cristalográfico das moléculas tinham impedido até agora o seu completo esclarecimento.

As principais conclusões deste estudo são:

a) a estrutura da MpgS de *T. thermophilus* HB27 foi resolvida, e é representativa da família das glicosiltransferases GT55, com um arranjo estrutural do tipo GT-A (www.cazy.org). Através das formas apo-MpgS e de um complexo com o substrato, [Mg^{2+}:GDP-α-D-manose], foram identificadas interacções-chave em dois laços flexíveis da enzima que contribuem para a ligação produtiva dos substratos. Um segundo ião metálico foi encontrado a 6 Å do grupo manosilo, e o seu papel catalítico ficou demonstrado através de estudos cinéticos e com mutantes de MpgS. Por outro lado, o resíduo Asp167 do motivo conservado D-X-D, foi considerado como potential nucleófilo catalítico à luz dos mecanismos actualmente propostos para as GTs retentoras da configuração anomérica nos seus produtos de reacção. A identificação do tipo de orientação glicosídica em relação ao motivo D/E-X-D, predominante nas GTs retentoras e nas inversoras da configuração anomérica, foi feita através de uma análise comparativa entre enzimas com arranjo estrutural GT-A com estruturas conhecidas contendo o nucleotídeo de

açúcar ou um seu análogo. Nas GTs retentoras, a ligação glicosídica está orientada para aquele motivo, enquanto nas GTs inversoras esta encontra-se orientada no sentido oposto. Esta característica ajuda a explicar a natureza estereoquímica das reacções, e constitui uma assinatura estrutural na distinção entre GTs retentoras e inversoras.

b) a elucidação estrutural da MpgP de *T. thermophilus* HB27, permitiu a sua classificação como uma fosfatase metalo-dependente pertencente à super-família das "Haloalcanoic Acid Dehalogenase-like phosphatase" (HAD-like), incluída numa família das fosfatases do tipo "Cof": a família das fosfatases de manosilo-3-fosfoglicerato. Esta estrutura é um excelente exemplo da combinação entre diferentes módulos funcionais numa nova família de enzimas com especificidade dirigida para novos substratos, mas preservando a base química da reacção. As formas estruturais "aberta" apo- e "fechada" holo- da MpgP estão relacionadas por rotação do módulo tipo "Cof" em torno de dois laços flexíveis da enzima, que ligam o módulo anterior ao módulo base "HAD-like". A dinâmica do centro catalítico foi estudada através de "instantâneos fotográficos cristalográficos", obtidos por crio-preservação de espécies representativas das várias etapas da coordenada da reacção. Os resultados sugerem que a hidrólise do grupo fosforilo do MPG ocorre através de um mecanismo concertado D_NA_N, onde o residuo Asp8 é o ácido catalítico para a formação do intermediário, sendo este subsequentemente atacado por uma molécula de água nucleofílica. Estas estruturas mostram aspectos mecanísticos chave na elucidação molecular da catálise de transferência de fosfomonoésteres que poderão ser extrapolados para outros membros da superfamília HAD. Genéricamente, estas características sugerem um *continuum* de mecanismos alternativos, que variam entre os que são exclusivamente dissociativos e os que são associativos.

c) Três estados conformacionais: aberto, semi-aberto e fechado foram identificados em *Eco*IDH por comparação de estruturas cristalográficas. Entre

outras, salientam-se a forma nativa *Eco*IDH, com um complexo *pseudo*-Michaelis-Menten e a forma mutante K100M, com os produtos da reacção. A conformação fechada é favorecida pela ligação ao substrato, a qual induz a rotação do domínio grande em torno da folha β-central em cada um dos monómeros do homodímero biológico. Resíduos catalíticos conservados ligam-se ao anel da nicotinamida e ao complexo do substrato, e movem-se como corpos rígidos durante o movimento de rotação. As conformações fechadas com o complexo *pseudo*-Michaelis-Menten da forma nativa e com o complexo ternário dos produtos da reacção, revelaram pela primeira vez uma imagem mais realista dos princípios mecanísticos da reacção de β-descarboxilação oxidativa. A análise das estruturas confirmou os papéis de ácido-base de Brønsted propostos para o par catalítico Tyr160-Lys230* anteriormente postulado.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF ABREVIATIONS

3D	Three-dimensional
3-PG	D-glycerate-3-phosphate
a.d.p.	atomic displacement parameters
A.U.	asymetric unit
A5P	Adenosine 2',5'-biphosphate
AKG	α-ketoglutarate
*Aper*IDH	*Aeropyrum pernix* IDH
β-PGM	*Lactococcus lactis* β-phosphoglucomutase
*Bsub*IDH	*Bacillus subtilis* IDH
D_N*$A_N$ss-like	dissociative-associative solvent separated nucleophilic substitution reaction like
DLS	Diamond Light Source
*Eco*IDH	*Escherichia coli* IDH
ESRF	European Synchrotron Radiation Facility
F.O.M	figure of merit
FBPase	*Sus scrofa* (Pig) fructose-1,6-bisphosphatase
GDP-Man	GDP-α-D-mannose
GH	*glycosyl hydrolases*
GpgS	glucosyl-3-phosphoglycerate synthase
GT	glycosyltransferase
HAD-like phosphatase	Haloalkanoid Acid Dehalogenase-like
HPP	*Bacteroides thetaiotaomicron* predicted hexose phosphate phosphatase
ICT	2R,3S-isocitrate
IDH	isocitrate dehydrogenase
IDH (K100M)	*Escherichia coli* IDH K100M mutant
IMDM	isopropylmalate dehydrogenase
MAD	Multiple-wavelength Anomalous Dispersion
MgS	Mannosylglycerate synthase
MM_{ca}	molecular mass
MPG	2-(α-D-mannosyl)-3-phosphoglycerate
MpgP	mannosyl-3-phosphoglycerate phosphatase
MPGP family family	Mannosyl-3-phosphoglycerate phosphatase
MpgS	mannosyl-3-phosphoglycerate synthase
$NADP^+$	β-Nicotinamide Adenine Dinucleotide Phosphate

NADPH	β-Nicotinamide Adenine Dinucleotide Phosphate reduced
NCS	non-crystallographic symmetry
NDPsugar	nucleoside-diphospho-sugar
NMP	ribosylnicotinamide-5'-phosphate
OXA	oxalosuccinate
PDB	Protein Data Bank
PO_3^- intermediate	metaphosphate intermediate
r.m.s	root-mean-square
R132H *Hsap*IDH1	R132H mutant of human IDH isoform 1
*Scer*IDH1	*Saccharomyces cerevisiae* IDH mitochondrial NADP-dependent
SLS	Swiss Light Source
S_N1-like reaction	nucleophilic substitution unimolecular -like
S_N2-like	nucleophilic substitution bimolecular-like reaction
S_Ni-like	nucleophilic substitution internal-like reaction
*Sscr*IDH	*Sus scrofa* (porcine) IDH
thio-$NADP^+$	β-thio-Nicotinamide Adenine Dinucleotide Phosphate
thio-NMP	ribosyl-thio-nicotinamide-5´-phosphate
Tm	melting temperature
TS	Transition state
w.r.t	with respect to
wtIDH	*Escherichia coli* IDH (wild type)
wtMpgS	wild-type MpgS

CHAPTER 1

INTRODUCTION

1.1. Outline on the current dynamics and structural aspects behind the catalytic power

Proteins play a myriad of roles such as: enzymatic catalysis, mechanical support, immune system protection, generation and transmission of nerve impulses. The understanding of protein structure-function relationships is one of the main tools for their rational classification into families and nested subfamilies.

Particularly, the role of protein structure in enzyme catalysis is well established, and conservation of structural features provides vital clues to their role in function (*1*). Proteins are inherently dynamical molecules that undergo structural fluctuations over a wide range of timescales, from femtoseconds to milliseconds or longer (*2, 3*). Conformational flexibility of enzymes has been associated with substrate (and cofactor) binding and product release, and is currently being topic of debate regarding to its relevance with the chemical steps and in the enzyme`s proficiency (*1, 2, 4*). In this view, flexible enzyme regions are found to be connected by conserved networks of coupled interactions that connect surface regions to active-site residues. In fact, different from the flexible structural regions, backbone flexibility profiles diverge slowly, being conserved both in protein family and superfamily (*5-7*).

A multidisciplinary approach is nowadays devoted toward a holistic view of enzyme structure flexibility and function (*1, 2*). Among other spectroscopic techniques, X-ray crystallography can provide snapshots of conformational enzyme states sampled during the reaction coordinate while NMR relaxation methods reveal the rates of interconversion between substrates and the corresponding relative populations (*8-10*).

Also, new models for computational analysis are being developed, in order to overcome the experimental limitations imposed by the time space where harmonic vibrations occur, and obtain insights into the molecular variables involved in the protein`s intramolecular signal transduction (*1, 3, 6, 11*).

However, these current experimental methods cannot simultaneously reveal the detailed atomic structures of the rare states and rationalize these findings with the intrinsic enzymes` motions that occur in a time scale as that of the catalytic turnover (*12*). With the advent of time-resolved crystallography (*13-16*), and with combined strategy of ambient-temperature X-ray crystallographic data collection, automated electron-density sampling and NMR analysis (*12*) steps are being made in this direction.

Other factors also play a role in enzyme catalysis, such as: the electrostatic effect of solvent reorganization; the active-site architecture which is set up to bind the substrate in a conformation approaching its transition state (TS), while at the same time provide a suitable environment for the stabilization of the TS species (1). In this context, the current view on the relationships between protein dynamics and function suggests that these are rooted in the free energy landscape and that fluctuations at equilibrium can influence biological functions (*6, 7, 17*).

The work presented hereafter comprises the structural elucidation of a glycosyltransferase, a phosphatase and oxidorreductase, by using X-ray crystallography diffraction methods. The conformational changes observed for each enzyme-substrate complex were compared to the corresponding unbound enzyme state for a detailed picture on the correlated dynamics in the catalytic power. Since X-ray crystallography is still one of the most accurate experimental methods for structure description, it will be used to rationalize these conformational changes as in terms of a series of correlated events that occur in the time scale of the reaction coordinate. Within this scope, slower protein movements occurring at microseconds (and longer; $>10^{-6}$ s) will be explored, as these correspond to global conformational fluctuations of large

domains, which include large displacements in surface loops, as well as, the coordinated movement of β-strands and α-helices by hinge motion. This will provide additional measurable data on the alternative enzymatic conformational states during the course of the reaction coordinate, and solidify the experimental background for its application into theoretical calculations on proteins` structural dynamics.

1.2 Molecular adaptations to extreme environments; compatible compounds in osmo- and thermostabilization

During these last 20 years, the interest in organisms that drive at extreme environmental conditions have caught the attention of many researchers, oriented towards uncovering the fundamental principles behind the strategies of adaptation and growth under such harsh conditions, as well as in finding novel resources tailored for biotechnological and medical applications.

Particularly, thermophiles have optimum growth temperatures between 65°C and 80°C, while hyperthermophilic organisms are those with optimum growth temperatures above 80°C (*18*). Many of these organisms are isolated from continental geothermal or artificial thermal environments, but some thermophiles have been isolated from marine hydrothermal environments, the best known of which are *Rhodothermus marinus and Thermus thermophilus* (*19, 20*). The water venting from continental hot springs is generally low in sodium and the isolates, being fresh water organisms, rarely grow in media containing more than 1.0% NaCl (w/v). Some organisms isolated from continental fresh water hot springs are, however, halotolerant with optimum growth in media without added NaCl but, like the strains of *Thermus thermophilus,* are able to grow in media containing 4.0-6.0% NaCl (*20*). On the other hand, hyperthermophilic organisms, are generally slightly halophilic, and optimum growth occurs in media containing 0.5 - 2.0% NaCl. Scientists interested in osmotic adjustment in (hyper)thermophiles are,

for now only able to examine halotolerant and slightly halophilic organisms (*21*).

One major finding brought from physiological cell response studies was the accumulation the small organic compounds, designated compatible solutes that preserve cell viability by balancing the fluctuations in the osmotic pressure of the external milieu (*22-24*). Initially classified as neutral or zwitterionic compounds, negatively charged compounds were identified in several (hyper)thermophiles, a feature which has been linked to their superior ability to act as protein stabilizers (*25, 26*). A wide variety of compounds is grouped into: amino acids and amino acid derivatives, sugars, sugar derivatives (heterosides) and polyols, betaines and the ectoines (*23*). Some are widespread in microorganisms, namely trehalose, glycine betaine and α-glutamates, while others are restricted to a few organisms. Polyols, for example, are widespread among *fungi* and *algae* but are very rare in bacteria and unknown in archaea. Ectoine and hydroxyectoine are examples of compatible solutes found only in bacteria (*21*). Storage of compatible solutes takes place by synthesis or by downregulation from the medium by means of special transport systems, activated by mechanical stimuli (*27*). Given that compatible solutes are implicated in defence against a variety of insults, e.g., too high or too low temperature, low water activity, high level of reactive species, high-dose of ionizing radiation, there is a demanding will for the elucidation of their protective role at the molecular level. This is particularly true in respect to “thermolecular” mechanisms underlying the stabilization of proteins by compatible solutes, a topic of great current interest as protein misfolding and aggregation is associated with a number of debilitating diseases, like Parkinson’s, Alzheimer’s, or Huntington’s (*28, 29*). Another pertinent question, is to know the specific effectiveness of each compound in the preservation/protection of the cell components. For example, while trehalose serves as a general stress protector, being implicated in a variety of stress responses (osmotic, heat, acid and oxidative stress), α-mannosylglycerate, a prototypical solute highly associated with

(hyper)thermophiles, is far better than trehalose for the preservation of protein structure against thermal denaturation (*25, 26, 30-33*). Understanding the nature of the molecular interactions, water/solute/protein, that promote the native fold, is a challenging issue that fuels vivid debate in the scientific community and demands powerful cross-disciplinary approaches (*33*).

Parallel to the urge in unveiling the molecular basis of thermo- and osmostabilization by these compounds, the knowledge on the osmosensing-regulatory mechanisms, and on the enzymology behind solute accumulation, are also far to be fully elucidated. With the advent of sequencing genomic projects, the number of putative genes involved in the biosynthesis or uptake of these compounds in response to stress is increasing at faster pace. This provides means for a clearer distribution of a particular set of genes as either being clustered within restricted phylogenetic branches, or alternatively, ubiquitously present among the three main Life Super-kingdoms (*33-35*). This comprises a “fingerprint” on the prevalence of a kind of strategy to cope with the environmental constrains and stress conditions. On the view of the applied research field, such distribution spectrum enlarges the scope of the applicability of these compounds, as well as, of potential biocatalysts with engineered superior catalytic efficiency to work under extreme reaction conditions. Thus, interest on the pathways for compatible solute biosynthesis is becoming current focus in the biotechnological industry.

A vast amount of data on the benefits and effectiveness of these compounds, as well as on their potential applications in the biotechnological industry have been reviewed elsewhere (*33, 35, 36*). Furthermore, the thermostability feature that accompanies mostly these enzymes opens new doors in the exploration of novel engineered enzymes, tailored for the specific needs of their industrial use.

Within the scope of the molecular mechanisms involved in the thermo- and osmoadataption response, herein we describe the structural elucidation of the two-step pathway of the compatible compound, α-mannosylglycerate in the thermophilic bacteria, *Thermus thermophilus* HB27.

1.3 Accumulation of α-mannosylglycerate in *Thermus thermophilus* HB27

The solute α-mannosylglycerate (MG), was initially identified in red *algae* of the order Ceramiales (*37*). More recently, this solute has been found in thermophiles, such as the bacteria *Rhodothermus marinus*, *Thermus thermophilus*, and *Rubrobacter xylanophilus*, and in many hyperthermophilic archaea (*38*). The effect of heat stress on the "solute pool" of *T. thermophilus* has not been studied, but knock-out mutants were used to demonstrate the role of MG during osmoadaptation of this extreme thermophile (*39*). Mannosylglycerate accumulates in *R. marinus* not only in response to osmotic stress, but also in response to supra-optimal growth temperature (*40*), by distinctively up-regulating one of the two pathways of MG biosynthesis (*41*) (Fig. 1.1).

The finding that MG is widespread in marine hyperthermophiles, and rarely found in mesophiles, led to the view that this osmolyte is involved not only in osmoprotection but also in thermoprotection of cellular components. In fact, MG possesses a superior ability to protect enzymes against thermal denaturation *in vitro,* and its potential usefulness in biotechnological and pharmaceutical applications has been often claimed (*26, 30*).

The genes and enzymes implied in the synthesis of MG have been characterised in several (hyper)thermophilic bacteria and archaea (*41-45*). In most organisms, MG synthesis proceeds via a two-step pathway: in the first step GDP-α-D-mannose and 3-phosphoglycerate are converted into mannosyl-3-phosphoglycerate by the action of mannosyl-3-phosphoglycerate synthase (MpgS; EC 2.4.1.217). The phosphorylated intermediate is subsequently dephosphorylated by mannosyl-3-phosphoglycerate phosphatase (MpgP; EC 3.1.3.70), yielding MG (Fig. 1.1). An alternative pathway has been biochemically characterized in *R. marinus* and comprises the direct transference of D-glycerate into GDP-α-D-mannose by mannosyl-glycerate synthase (MgS; EC 2.4.1.269) (*41, 44*).

Inspection of public databases resulted in the identification of MPGS homologues in the mesophilic bacteria *Dehalococcoides ethenogenes*, and *Verrucomicrobiae* bacterium, and also in several fungi. The MPGS activity of the respective gene products has been confirmed in *Neurospora crassa* and *Dehalococcoides ethenogenes* (*46, 47*). Therefore, the synthesis of mannosylglycerate seems not restricted to (hyper)thermophiles, but its physiological role in algae is unknown, and the accumulation of this solute in other mesophiles awaits demonstration. It is expected that the investigation of the molecular evolution of mannosylglycerate synthesis will shed light on the origin and dissemination of this trait in the Tree of Life (*33*).

Biochemical characterization of the enzyme led to the classification of the MpgS in the retaining Glycosyltransferase family 55 (GT55; www.cazy.org). Thus far, this family comprises only proteins homologous to *T. thermophilus* MpgS, whereas MgS is classified into the retaining GT78 family, and a "front-face" S_Ni-like mechanism for glycosyl transfer with retention of anomeric configuration was proposed[1].

However, the catalytic mechanisms of MpgS and MpgP enzymes, were not been investigated at the beginning of this thesis, and were taken part of the global study on the molecular determinants that dictate the fundamental mechanisms of enzymatic catalysis.

---

[1] The crystallographic structure elucidation of MgS from *R. marinus*, was carried out by our team, at the time as the same structure was being published by Flint and co-workers (*48*). No major crystallographic differences were observed.

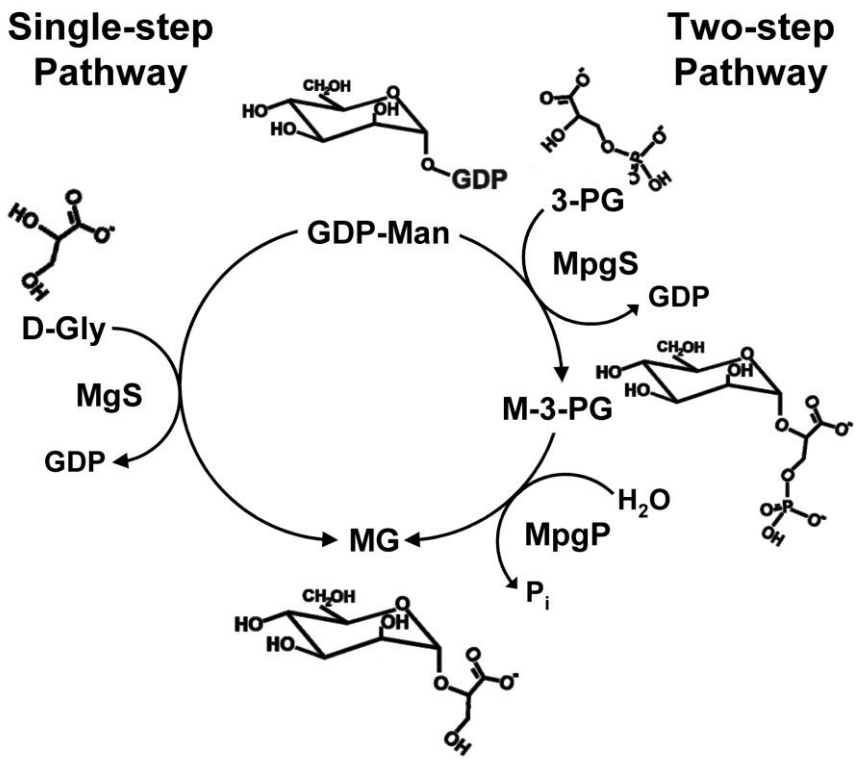


Figure 1.1 The two pathways for the synthesis of mannosylglycerate. Two-step pathway (right): in the first step, MpgS catalyses the transfer of the mannosyl moiety from GDP-Man into 3-PG, yielding MPG; in the second step, this intermediate compound is hydrolyzed by MpgP to form the final product α-mannosylglycerate (MG). Single-step pathway (left): MgS catalyses the direct glycosyl-transfer of GDP-Man into D-Glycerate (D-Gly). GDP-Man, GDP-α-D-mannose; 3-PG, 3-D-phosphoglycerate, D-Gly, D-glycerate; MG, α-mannosylglycerate; MpgS, mannosyl-3-phosphoglycerate synthase; MpgP, mannosyl-3-phosphoglycerate phosphatase; MgS, mannosylglycerate synthase.

1.4 Glycosyltransferases; fold, mechanism and biological implications.

Oligosaccharides and glycoconjugates are functional building blocks with a wide array of essential biological functions ranging from energy storage and utilization to complex information display systems that modulate normal cell function (49). The mechanistic strategies used to make and break this class of covalent bond (ΔG ~30 kcal/mol) with a high degree of both *stereo*- and *regio*-selectivity is a topic of considerable interest in various fields of chemistry and chemical biology (49). The enzymes responsible such diversity comprise the glycosidases, glycosyltransferases and phosphorylases (Fig. 1.2). The most notorious example lies in all of domains of life; the biosynthesis of complex glycans requires the concerted action of a multitude of Glycosyltransferases enzymes (GTs; EC 2.4.x.y), that catalyze the transfer of a mono- or oligosaccharide from a glycosyl donor (for example, a sugar-nucleotide) to a suitable acceptor [(for example, a glycan, peptide or lipid (*50*)]. Particularly, the role of GTs extents trough many fundamental biological processes underpinning human health and disease, such as cell signalling, cellular adhesion, carcinogenesis and cell wall biosynthesis in human pathogens (*50*).

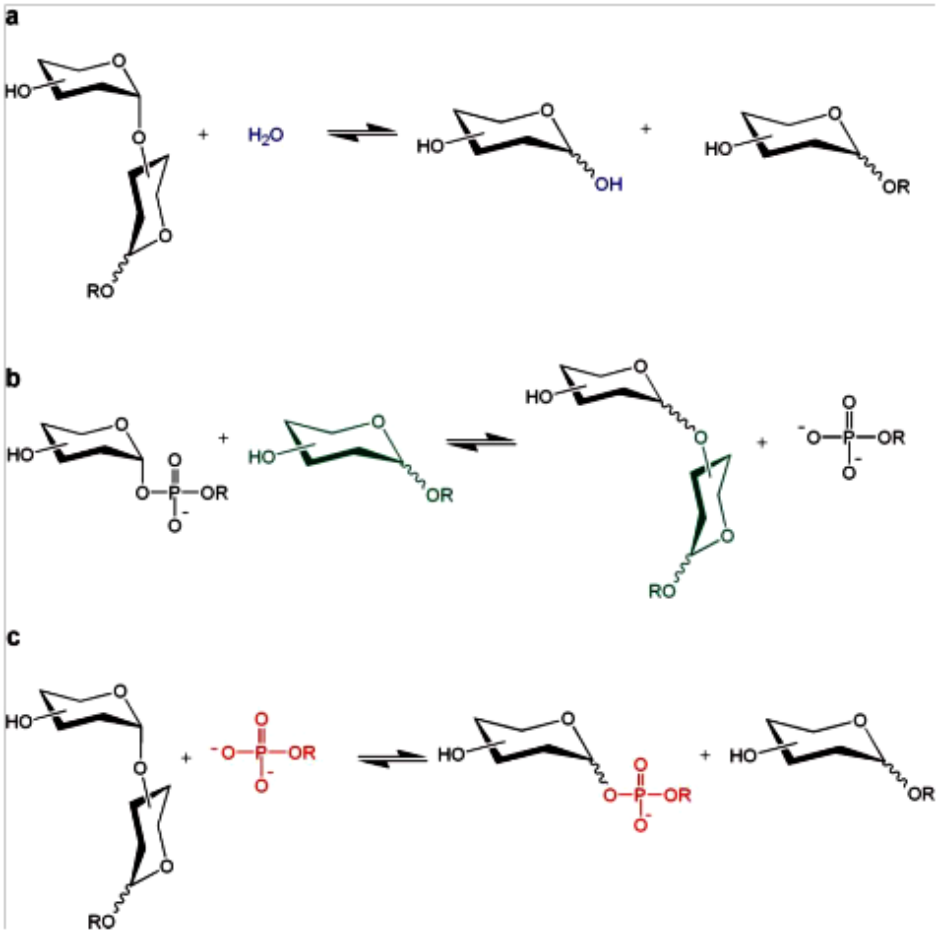


Figure 1.2 Overall reactions catalyzed by (a) glycosidases, (b) glycosyltransferases, and (c) phosphorylases (*49*).

Sequence-based classification spreads GTs in many families by amino acid sequence similarities, [(http://afmb.cnrs-mrs.fr/CAZY) (*51*)], thus reflecting the variety of molecules that can be used as acceptors. In addition to nucleotide-diphospho sugar-donor (NDP-sugar) dependent enzymes, they also utilize dolichol-phospho-sugars, sugar-1-phosphates and lipid diphospho-sugars as activated donors.

Extending the same rational as that applied for the GHs (*52*) for such classification scheme one would expect to integrate both structural and

mechanistic features within the same family. Yet, the promiscuity for either an inverting or retaining [2] mechanism present within the same family is becoming a current feature, and thus novel approaches for the rational classification of the GTs are mandatory .

In contrast to GHs that adopt a large variety of folds, including all α all β, or mixed α/β structures, GTs folds have been observed to consist primarily of α/β/α sandwiches similar or very close to the Rossmann type fold, a classical structural motif (six-stranded parallel β-sheet with 321456 topology) found in many nucleotide binding proteins (*53*). Until recently only two structural superfamilies have been described for GTs, named GT-A and GT-B (Fig. 1.3), and which were first observed in the original Spore coat polysaccharide biosynthesis protein and β-glucosyltransferase structures, respectively (*54, 55*). A third family has recently emerged which comprises the bacterial sialyltransferase (CstII) belonging to family GT42 [based on the β-strand connectivity (*56*)]. GT-A and GT-B folds are also shared by non-GT enzymes, such as nucleotidyltransferases and sugar epimerases, respectively. The GT-B fold consists of two separate Rossmann domains with a connecting linker region and a catalytic site located between the domains. There is an excellent structural conservation between protein members of the GT-B family, particularly in the C-terminal domain which corresponds to the nucleotide-binding domain. Variations are more pronounced in the N-terminal domains, in the loops and helices which point towards the active site, which have evolved to accommodate very different acceptors. The structural features of GT-A fold enzymes will be herein focused, whereas an extended structural survey describing GT-B fold enzyme features can be found elsewhere (*57, 58*).

The GT-A fold consists of an α/β/α sandwich (a seven stranded β-sheet with 3214657 topology in which strand 6 is antiparallel to the rest) that resembles a Rossmann fold. The central β-sheet is flanked by a smaller one, and the association of both creates the active site (Fig. 1.3). A general feature of all enzymes of the GT-A family is the presence of a common motif, the DxD motif, and also their requirement for a divalent cation for activity (*57, 58*). The

DxD motif is shown in all crystal structures to interact primarily with the phosphate groups of nucleotide donor through the coordination of a divalent cation, typically Mn^{2+} (*58*). Comparison of the catalytic domains of enzymes of the GT-A family revealed the presence of two regions that are structurally well conserved in all members of the GT-A family, including inverting and retaining enzymes. This suggests that common structural elements are necessary for the glycosyl transfer reaction, irrespective of the stereochemistry of the reaction. The first region mostly corresponds to the Rossmann-type nucleotide-binding domain, encompassing the first 100-120 residues, and that is terminated by the DxD motif. The key amino acids that interact with NDP are mainly found at the C-terminal of strands β1 and β4. In some crystal structures, residues in the C-terminus of the catalytic domain were shown to make additional contacts with the NDP moiety.

While the mechanistic strategies used by GlycosylHydrolases (GHs), to catalyse glycosidic bond hydrolysis are fairly well understood on both a structural and chemical level (*61, 62*), the mechanistic understanding of the GTs responsible for glycoside bond formation has lagged far behind. Despite a lack of evolutionary relatedness, by simple chemical analogy, GTs are thought to use mechanistic strategies that directly parallel those used by GHs and transglycosidases (*63, 64*).

---

[2] By analogy with glycosidases, two main stereochemical outcomes exist for glycosyltransferases: inversion of the anomeric configuration (for instance UDP-glucose -> β-glucoside) or retention of the anomeric configuration (for instance UDP-glucose -> α-glucoside).

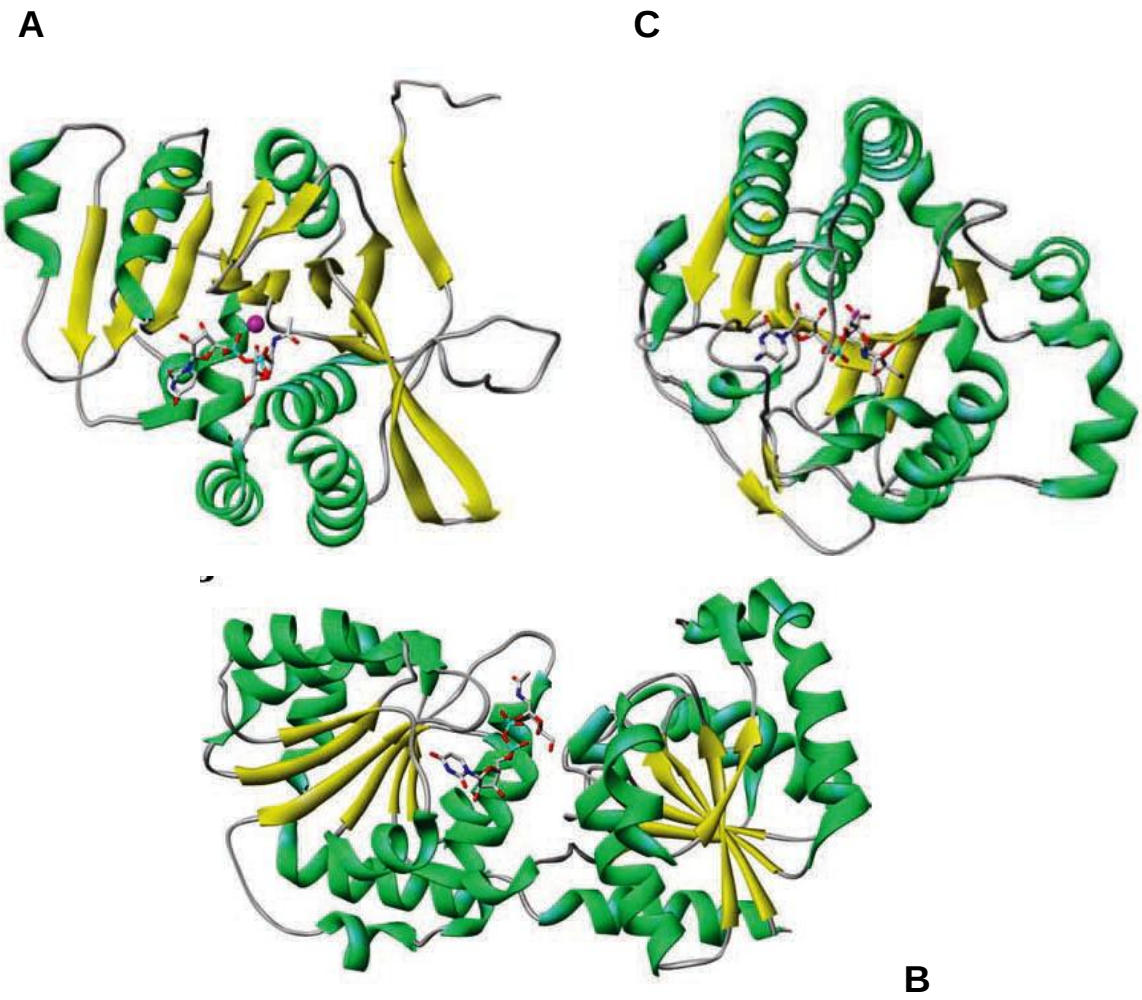


Figure 1.3 Ribbon diagram of three glycosyltransferases (GTs) representative of the different folds. Bound nucleotide sugar are represented with stick model, and manganese, when present, by a ball. (**A**) GT-A fold, mouse α-1,4-*N*-acetylhexosaminyltransferase (EXTL2) complexed with UDP-Gal-NAc (PDB code 1OMZ) (*59*), (**B**) GT-B fold, *Escherichia coli* MurG complexed with UDP-GlcNAc (PDB code 1NLM) (*60*), and (**C**) *Campylobacter jejuni* sialyltransferase CstII complexed with cytidine monophospho 3-fluoro N-acetyl neuraminic acid (CMP-3FNeuAc) (PDB code 1RO7) (*56*). Figure obtained from reference (*57*).

Nucleophilic substitution at the anomeric carbon of the transferred glycosyl residue proceeds with either inversion or retention of configuration of the donor substrate and each stereochemical course necessitates a distinct catalytic mechanism (Fig. 1.4). It is currently not clear how stereochemical control in glycosyltransferase-catalyzed group transfer is achieved. In fact, one of the fundamental challenges in the field of glycobiology remains the dissection of the catalytic mechanism of nucleotide-sugar dependent GTs, especially those that act with retention of anomeric configuration (*65*).

Inverting GTs are proposed to utilize a direct-displacement S_N2-like reaction mechanism in which departure of the leaving group is facilitated by a Lewis acid, and nucleophilic attack is assisted by a catalytic base of the enzyme. Drawing analogy to GHs for which reaction coordinates leading to inversion or retention have been very well-characterized (*63*) a double displacement-like reaction involving a covalent glycosyl-enzyme intermediate appeared to be the preferred choice of mechanism for the retaining GTs. However, GTs crystal structures revealed a surprising lack of conserved architecture in the region of the active site where the catalytic nucleophile would have to be positioned (*58*). In addition, despite exhaustive biochemical studies with techniques that have been successfully used in the characterization of retaining glycoside hydrolases (*66, 67*) evidence supporting a covalent intermediate for retaining GTs has remained elusive. Major barrier to the study of these enzymes is the lack of non-hydrolysable substrates or fluoro-sugar type inhibitors with affinities similar to that of the substrates, has rendered ineffective the ability in altering the relative rates of glycosylation versus deglycosylation steps, thus in trapping intermediate species on retaining GTs (*67*), and in allowing the structural access to the ternary complex or something that resembles it.

Glycosyltransferase inhibitors are rare; rarer still are those which do not harness portions of the nucleotide donor (*68*). Only one bisubstrate analogue of a retaining GT has previously been described, and surprising inhibition profiles were observed, inconsistent with a double-displacement mechanism

(*69*). Partially as a consequence of this lack of suitable compounds, there are no 3-D structures of intact ternary complexes with which to describe the catalytic centre and reveal geometry of the transfer process (*65*). An alternative mechanism, often referred to as internal return S_Ni-like (IUPAC defenition $D_N{}^*A_{Nss}$), was therefore considered. Chemical precedent for this type of a mechanism comes from detailed kinetic and conformational studies of the solvolysis of glucose derivatives in mixtures of ethanol and trifluoroethanol (*70*). The S_Ni-like mechanism has since been proposed for other structurally defined retaining transferases based on the lack of appropriately positioned nucleophiles in their active sites (*59, 71*) and had previously been proposed for the structurally similar glycogen phosphorylase (*72*). This is proposed to involve a short-lived ion pair intermediate, the formation of which arguably requires electrostatic front-side stabilization from the departing (substituted) phosphate and perhaps a certain amount of nucleophilic “push” from an enzyme group positioned on the backside of the glycosyl ring. In addition to stabilizing the intermediate, hydrogen bonding between the incoming nucleophile and the leaving group would play an important base catalytic role during activation of the acceptor hydroxyl group for nucleophilic attack.

The study and comparison of the enzymes responsible for making and breaking glycosidic linkages provides insight into the nature of mechanistic evolution. A direct analogy between the straightforward mechanisms of inverting glycosidases and glycosyl transferases is apparent. However, when it comes to a comparison between the mechanisms of retaining enzymes, while recent insights may serve to provide a key, it has not yet been turned to unlock the door and a distinct relation remains unclear (*49, 58*).

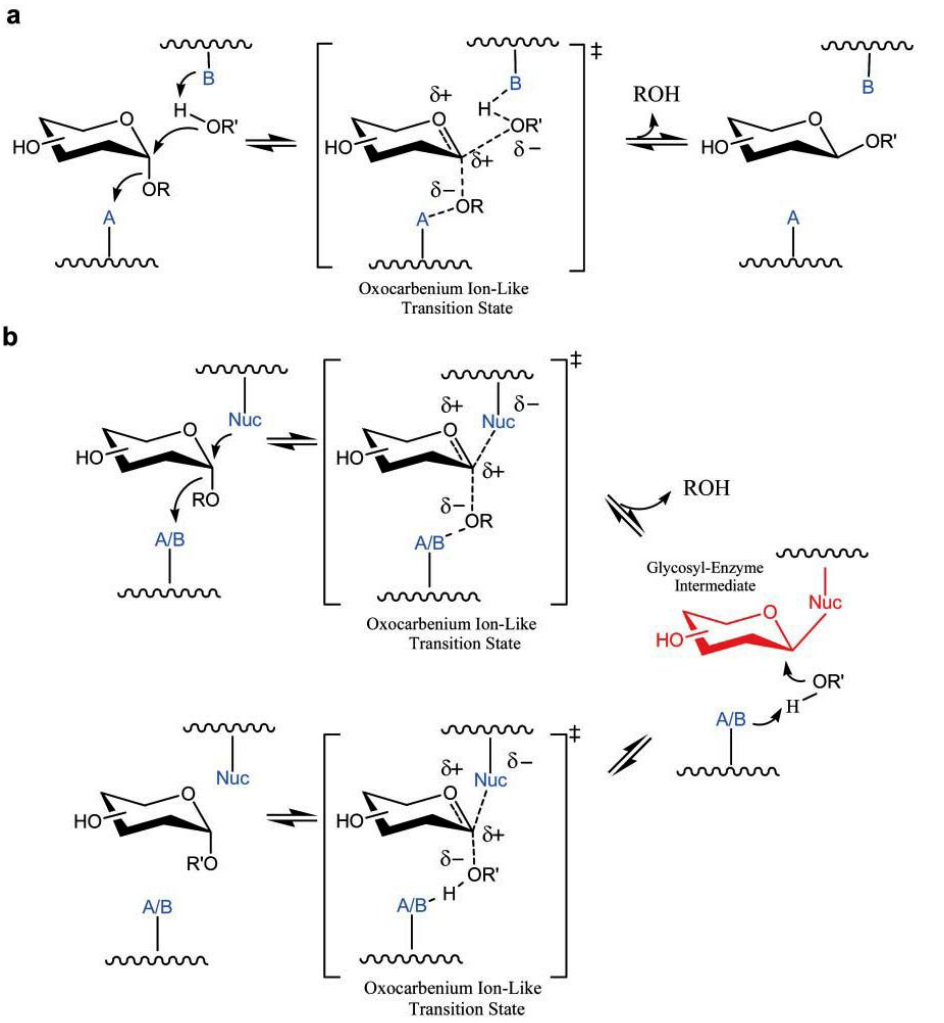

a
B
H
OR'
HO
O
OR
A
δ+
δ−
‡
ROH
Oxocarbenium Ion-Like
Transition State
b
Nuc
RO
A/B
R'O
Glycosyl-Enzyme
Intermediate
c
NH2
R
H—OR'

Figure 1.4 Established and proposed mechanisms for glycosidases, transglycosidases and glycosyltransferases, respectively. For glycosidases, R = a carbohydrate derivative and R'OH =H_2O or phosphate (phosphorylases classified as glycosidases). For transglycosidases, R = a carbohydrate derivative and R'OH = another sugar. For glycosyltransferases, R = a nucleoside diphosphate (e.g. UDP, GDP), a lipid phosphate, or phosphate (phosphorylases classified as glycosyltransferases) and R'OH = an acceptor group (e.g. another sugar or a protein). (**a**) A direct displacement S_N2-like reaction results in inverted anomeric configuration via a single oxocarbenium ion-like transition state. (**b**) The double displacement mechanism proceeds via two oxocarbenium ion like transition states with the intermediate formation of a discrete covalently bound glycosyl-enzyme species, resulting in overall retention of anomeric configuration. (**c**) "S_Ni-like" mechanism proposed for retaining glycosyltransferases and glycogen phosphorylase involving a direct "front-side" displacement resulting in retention of anomeric configuration. Scheme reaction obtained from (*49*).

1.5 Mechanistic alternatives in phosphate monoester transfer catalysis.

The hydrolysis of phosphate esters is an important reaction in many biological systems. DNA and RNA are phosphate diesters, while many intermediates in the metabolism exist as phosphate monoesters. Phosphorylation of proteins is an important control mechanism. While triesters are not naturally occurring biological molecules, enzymes have evolved to hydrolyse these man-made toxic compounds. However, monoesters and diesters have well-known roles in genetic materials, in coenzymes and in energy reservoirs, and as intermediates in biological transformations, while monoesters formed by protein phosphorylation have key roles in the regulation of host processes (*73*).

Phosphoryl transfer reactions are substitution reactions at phosphorus, and three limiting mechanisms exist (Fig. 1.5). One is a dissociative, S_N1-type mechanism ($D_N + A_N$ in the IUPAC nomenclature). In such a mechanism, a stable metaphosphate ion (PO_3^-) is attacked by a nucleophile in a subsequent rate-determining step. Other phosphoryl transfer mechanisms are an associative, two-step addition-elimination mechanism ($A_N + D_N$) with a phosphorane intermediate, and a concerted mechanism (A_ND_N) with no intermediate. In the concerted A_ND_N mechanism, bond formation to the nucleophile and bond fission to the leaving group both occur in the transition state. The prevailing chemical pathway is determined by the nature of the nucleophile, electrophile and leaving group and by the solvent (*74-76*).

Phosphotransferases, the enzymes that catalyze phosphoryl-transfer reactions, enhance the reaction rate by up to ~10^{21} by lowering the energy barrier for the reaction via strong binding to the transition state [K_d ~10^{-26} M (*77*)]. The currently dominating view regarding phosphate ester hydrolysis in solution is that diesters and triesters follow a more associative-like pathway, whereas the hydrolysis of monoesters proceeds by a dissociative mechanism. The latter does not necessarily mean that free metaphosphate is formed as an intermediate, but it is rather interpreted such that the rate-limiting TS mainly involves leaving-group bond fission with little bond formation to the incoming

nucleophile (*73, 78*). It would then be expected that the total axial phosphorous-oxygen bond order is lower in the TS than in the reactants and products (*79, 80*). Metaphosphate has been observed in the X-ray structure of fructose-1,6-bisphosphatase grown in an equilibrium mixture of substrate and product (*81*), and a moiety that could be described as a stabilized metaphosphate has been observed in the structure of α-phosphoglucomutase from *Lactococcus lactis* obtained at cryogenic temperature (*82*). This dogma is, however, repeatedly challenged by enzyme studies in which phosphate monoester cleavage is often interpreted in terms of more or less associative bimolecular mechanisms (*73, 83, 84*). It is also possible that the energetics of the associative and dissociative mechanisms are similar in solution and that enzymes catalyzing phosphate monoester (P-O bond) cleavage actually can alter the mechanism of their uncatalyzed reaction counterparts (*85*). What then controls the chemical pathway of the enzymes that catalyze phosphoryl-transfer reaction?

Chapter 3 describes the structural elucidation of *T. thermophilus* HB27 MpgP, a metal-dependent Haloalcanoic Acid Dehalogenase-like phosphatase (HAD-like), and provides principle mechanistic features for a concerted D_NA_N phosphoryl transfer mechanism, a paradigm which can be applied to other members of the HAD-like superfamily.

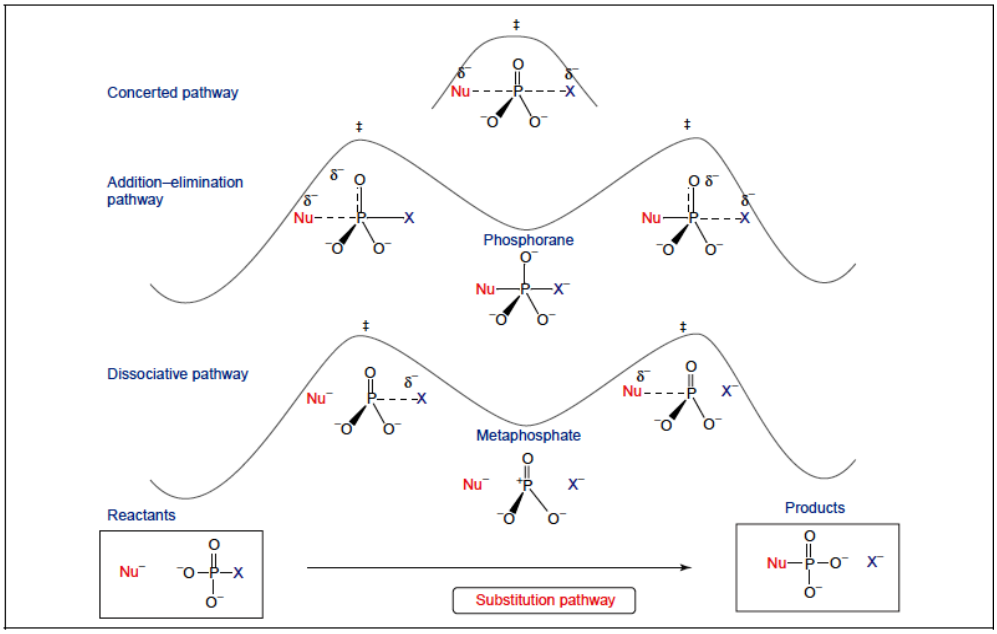


Figure 1.5 Three possible limiting chemical mechanisms for phosphoryl transfer. The transfer between two nucleophiles (Nu) differs in the timing of bond formation and cleavage. A hypothetical free energy diagram for each path is shown in the background. These highlight the fact that in the case of the addition-elimination (A_N + D_N; associative) and dissociative mechanisms (D_N + A_N) an intermediate is formed, whereas there is no intermediate in the concerted reaction (D_NA_N). In the addition-elimination mechanism, bond making to form a pentavalent phosphorane intermediate precedes bond breaking. In the dissociative mechanism, bond breaking to form a metaphosphate intermediate precedes bond making. In the concerted reaction, bonds are simultaneously made and broken in a single step, and the transition state resembles either metaphosphate or phosphorane intermediates. Scheme reaction obtained from(*86*).

1.6 Isocitrate dehydrogenase: a paradigm for a concerted structural dynamics with catalytic proficiency

Isocitrate dehydrogenase [IDH; 2R, 3S-isocitrate:$NADP^+$ oxidoreductase (decarboxylating); EC 1.1.1.42] belongs to a large family of pyridine nucleotide-linked β-hydroxyacid decarboxylating dehydrogenases. IDH converts isocitrate into α-ketoglutarate, and is the first enzyme in the CO_2 - evolving steps of the tricarboxylic acid cycle with concomitant reduction of $NAD(P)^+$ to NAD(P)H. The overall reaction catalyzed by IDH involves two steps: the activation of isocitrate into its oxaloacetate intermediate and β-decarboxylation followed by hydrogenation yielding α-ketoglutarate (Fig. 1.6).

Eukaryotic cells express two distinct IDH classes: the NAD^+-dependent enzyme (EC 1.1.1.41) found only in mitochondria and displaying allosteric properties, and a non-allosteric $NADP^+$-dependent enzyme (EC 1.1.1.42) that is found in both mitochondria and cytoplasm. In humans, the impaired ability of $NADP^+$-dependent IDHs (IDH1 and IDH2 isoforms) to convert isocitrate into α-ketoglutarate is known to be related with metabolic disorders having severe phenotypic outcomes, such as the cancer-associated IDH1 malignant progression of gliomas (*87, 88*). In *Escherichia coli,* as well as other bacterial cells, only a single $NADP^+$-dependent IDH is found (*89*) and it is a key regulator in the bypass of the TCA cycle to the glyoxylate pathway when acetate is used as carbon source (*90, 91*). The IDH regulatory activity mechanisms are thus of major interest as they act as checkpoints for modulating cell homeostasis.

The homodimeric IDH topology is composed of a central β-sheet that spans along the large α+β region and supersecondary α/β region the latter distinctive from the canonical Rossman-like fold domain that prevails in members related to the lactate dehydrogenase family (LDH family) (*92, 93*). Despite of a topology markedly different from that of the LDH fold, $NADP^+$ might conceivably bind near the N-termini of one or more helices in a manner analogous to NAD^+ binding to LDH (*94*). The coenzyme-binding site confirms the fundamentally different evolution of the isocitrate dehydrogenase-like and

the lactate dehydrogenase-like classes of nucleotide-binding protein, thus a potential alternative solution for the same evolutionary challenge (93, 94). This has led to the inclusion of the prior into the dimeric isocitrate/isopropylmalate dehydrogenase family (SCOP accession 53660).

IDH activation and inactivation in *E. coli* is controlled by a unique bifunctional regulatory enzyme, the IDH kinase/phosphatase AceK (EC 2.7.11.5), via “on-off switch” mechanism as the phosphorylated form of *Eco*IDH at Ser113 residue has no activity (96-98).

The AceK-IDH complex has long been considered a prototypic model system for protein phosphorylation in prokaryotes (*96, 99*) and recently, the structure of the complex AceK-IDH of *E. coli* has been determined shedding light into the molecular basis for AceK multifunctionality. The AMP-mediated conformational changes act as a switch between the kinase and phosphatase activities, in addition with the high order of interaction and recognition between AceK and IDH(*99*). On the other hand, eukaryotic $NADP^+$-linked IDHs rely, in a self-regulatory based electrostatic repulsive mechanism that mimics the prokaryotic regulation by phosphorylation (*89, 100*) while NAD^+-linked IDHs are allosterically regulated (*101*).

Structural studies in *Eco*IDH provided pioneering resource data on the correlated structural dynamics with the catalytic proficiency (*13, 102, 103*), in an analogous fashion to that proposed by isopropylmalate dehydrogenase (*95*). A paradigm is built; the catalytic competent active site is engaged in concert the binding of $NADP^+$ and substrate:metal complex. Also, dissection of the structural determinants involved in the cascade of electrostatic interactions provided principles for the molecular basis of substrate-dependent structural dynamics characterized by hinge-closure motion from ‘open’ (apo) to ‘closed’ (holo) conformations. Yet, a complete picture of a catalytic competent reactive center was yet to be obtained from *Eco*IDH crystals. Parallel crystallographic studies with homologous IDHs gave novel insights into the catalytic regulation of the homodimeric IDHs and in the molecular mechanisms of β-oxidative decarboxylation reaction (*87, 104, 105*). Though, fundamental questions

remained still to be answered, such as the identities of the assisting Brönsted acid/base catalyst.

Chapter 4 describes novel structures of *Eco*IDH that mirror the dehydrogenation and β-decarboxylation steps of the reaction cycle. Together with the subtract-induced protein dynamics, details on the chemistry behind the β-decarboxylation reaction are explained in light of our structural data and which can be generalised to all homodimeric $NADP^+$-linked IDHs.

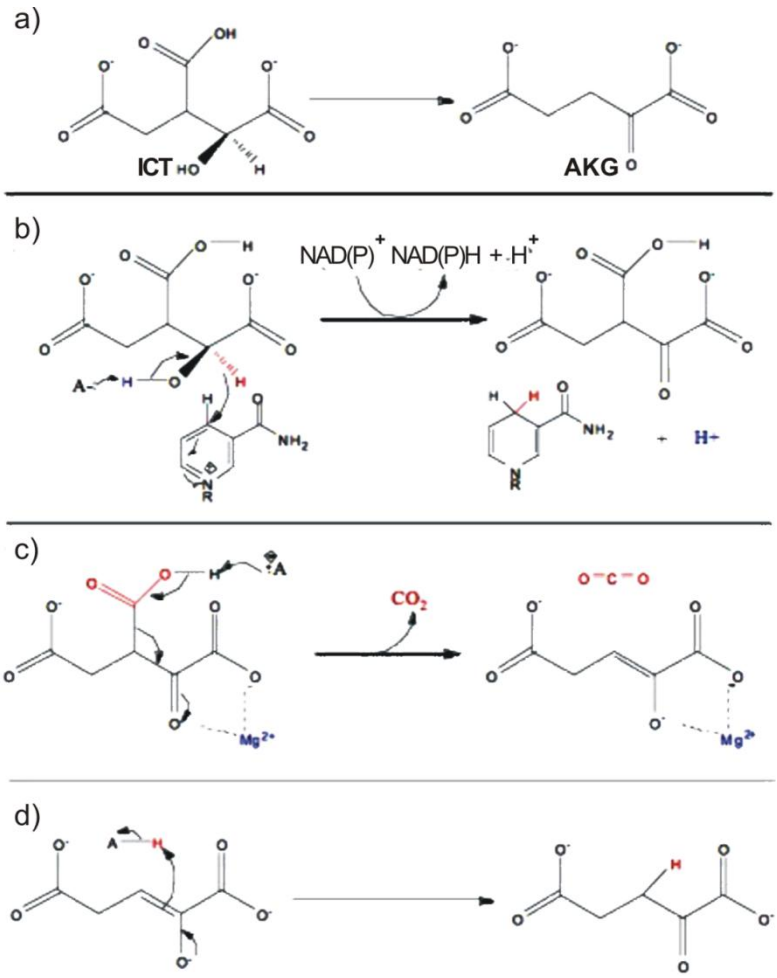


Figure 1.6 Isocitrate dehydrogenase (IDH) catalytic mechanism a) IDH converts isocitrate (ICT) into α-ketoglutarate (AKG) by a two-step β-oxidative decarboxylation reaction. **b)** the first step of the reaction involves the oxidation of ICT, to the intermediate oxalosuccinate (OXA) with C2 hydride transfer to the *re* face of the nicotinamide C4 of $NAD(P)^+$, **c)** second step comprises facile decarboxylation at C3 followed by hydrogenation, yielding AKG as in **d)**. The dias A^-/AH corresponds to the Brönsted assisting base/acid pair catalysts. Picture obtained from Wikipedia.

REFERENCES

1. Ramanathan, A., and Agarwal, P. K. (2011) Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis, *PLoS Biol 9*, e1001193.
2. Agarwal, P. K. (2006) Enzymes: An integrated view of structure, dynamics and function, *Microb Cell Fact 5*, 2.
3. Ma, C. W., Xiu, Z. L., and Zeng, A. P. (2011) A new concept to reveal protein dynamics based on energy dissipation, *PLoS One 6*, e26453.
4. Schwartz, S. D., and Schramm, V. L. (2009) Enzymatic transition states and dynamic motion in barrier crossing, *Nat Chem Biol 5*, 551-558.
5. Law, A. B., Fuentes, E. J., and Lee, A. L. (2009) Conservation of side-chain dynamics within a protein family, *J Am Chem Soc 131*, 6322-6323.
6. Papaleo, E., Tiberti, M., Invernizzi, G., Pasi, M., and Ranzani, V. (2011) Molecular determinants of enzyme cold adaptation: comparative structural and computational studies of cold- and warm-adapted enzymes, *Curr Protein Pept Sci 12*, 657-683.
7. Villali, J., and Kern, D. (2010) Choreographing an enzyme's dance, *Curr Opin Chem Biol 14*, 636-643.
8. Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins, *Nature 450*, 964-972.
9. Mittermaier, A., and Kay, L. E. (2006) New tools provide new insights in NMR studies of protein dynamics, *Science 312*, 224-228.
10. Teodoro, M. L., Phillips, G. N., Jr., and Kavraki, L. E. (2003) Understanding protein flexibility through dimensionality reduction, *J Comput Biol 10*, 617-634.
11. Daniel, R. M., Dunn, R. V., Finney, J. L., and Smith, J. C. (2003) The role of dynamics in enzyme activity, *Annu Rev Biophys Biomol Struct 32*, 69-92.

12. Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., and Alber, T. (2009) Hidden alternative structures of proline isomerase essential for catalysis, *Nature 462*, 669-673.
13. Bolduc, J. M., Dyer, D. H., Scott, W. G., Singer, P., Sweet, R. M., Koshland, D. E., Jr., and Stoddard, B. L. (1995) Mutagenesis and Laue structures of enzyme intermediates: isocitrate dehydrogenase, *Science 268*, 1312-1318.
14. Cho, H. S., Dashdorj, N., Schotte, F., Graber, T., Henning, R., and Anfinrud, P. (2010) Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering, *Proc Natl Acad Sci U S A 107*, 7281-7286.
15. Genick, U. K., Borgstahl, G. E., Ng, K., Ren, Z., Pradervand, C., Burke, P. M., Srajer, V., Teng, T. Y., Schildkamp, W., McRee, D. E., Moffat, K., and Getzoff, E. D. (1997) Structure of a protein photocycle intermediate by millisecond time-resolved crystallography, *Science 275*, 1471-1475.
16. Moffat, K. (1989) Time-resolved macromolecular crystallography, *Annu Rev Biophys Biophys Chem 18*, 309-332.
17. Nashine, V. C., Hammes-Schiffer, S., and Benkovic, S. J. (2010) Coupled motions in enzyme catalysis, *Curr Opin Chem Biol 14*, 644-651.
18. Blöchl, E., Burggraf, S., Fiala, G., Lauerer, G., Huber, G., and Huber, R. (1995) Isolation, taxonomy and phylogeny of hyperthermophilic microorganisms, *World J Microbiol Biotechnol 11*, 9-16.
19. Alfredsson, G. A., Kristjánsson, J. K., Hjörleifsdottir, S., and Stetter, K. O. (1988) Rhodothermus marinus, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland, *J Gen Microbiol 134*, 299-306.
20. da Costa, M. S., Nobre, M. F., and Rainey, F. A. (2001) The genus *Thermus*, in *Bergey's Manual of Systematic Bacteriology.* (Boone, D. R. a. C., R.W., Ed.) 2nd ed., pp 401-414, Springer, New York.

21. Santos, H., and da Costa, M. S. (2002) Compatible solutes of organisms that live in hot saline environments, *Environ Microbiol 4*, 501-509.
22. Brown, A. D. (1976) Microbial water stress, *Bacteriol Rev 40*, 803-846.
23. da Costa, M. S., Santos, H., and Galinski, E. A. (1998) An overview of the role and diversity of compatible solutes in Bacteria and Archaea, *Adv Biochem Eng/Biotechnol 61*, 117-153.
24. Roberts, M. F. (2005) Organic compatible solutes of halotolerant and halophilic microorganisms, *Saline Systems 1*, 5.
25. Faria, T. Q., Knapp, S., Ladenstein, R., Macanita, A. L., and Santos, H. (2003) Protein stabilisation by compatible solutes: effect of mannosylglycerate on unfolding thermodynamics and activity of ribonuclease A, *ChemBioChem 4*, 734-741.
26. Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C., and Santos, H. (2008) Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms, *Carbohydr Res 343*, 3025-3033.
27. Sochocka, M., and Boratynski, J. (2011) [Osmoregulation--an important parameter of bacterial growth], *Postepy Hig Med Dosw (Online) 65*, 714-724.
28. Lee, V. M., and Trojanowski, J. Q. (2006) Mechanisms of Parkinson's disease linked to pathological alpha-synuclein: new targets for drug discovery, *Neuron 52*, 33-38.
29. Skovronsky, D. M., Lee, V. M., and Trojanowski, J. Q. (2006) Neurodegenerative diseases: new concepts of pathogenesis and their therapeutic implications, *Annu Rev Pathol 1*, 151-170.
30. Borges, N., Ramos, A., Raven, N. D., Sharp, R. J., and Santos, H. (2002) Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes, *Extremophiles 6*, 209-216.

31. Faria, T. Q., Lima, J. C., Bastos, M., Macanita, A. L., and Santos, H. (2004) Protein stabilization by osmolytes from hyperthermophiles: effect of mannosylglycerate on the thermal unfolding of recombinant nuclease a from *Staphylococcus aureus* studied by picosecond time-resolved fluorescence and calorimetry, *J Biol Chem 279*, 48680-48691.
32. Ramos, A., Raven, N., Sharp, R. J., Bartolucci, S., Rossi, M., Cannio, R., Lebbink, J., Van Der Oost, J., De Vos, W. M., and Santos, H. (1997) Stabilization of Enzymes against Thermal Stress and Freeze-Drying by Mannosylglycerate, *Appl Environ Microbiol 63*, 4020-4025.
33. Santos, H., Lamosa, P., Borges, N., Gonçalves, L. G., Pais, T. M., and M. V. Rodrigues, M. V. (2011) Organic Compatible Solutes of Prokaryotes that Thrive in Hot Environments: The Importance of Ionic Compounds for Thermostabilization, in *Extremophiles Handbook* (Horikoshi, K., Antranikian, G., Bull, A. T., Robb, F. T., and Stetter, K. O., Eds.), pp 497-520, Springer, Tokyo.
34. Empadinhas, N., and da Costa, M. S. (2008) To be or not to be a compatible solute: bioversatility of mannosylglycerate and glucosylglycerate, *Syst Appl Microbiol 31*, 159-168.
35. Empadinhas, N., and da Costa, M. S. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea, *Environ Microbiol 13*, 2056-2077.
36. Luley-Goedl, C., and Nidetzky, B. (2011) Glycosides as compatible solutes: biosynthesis and applications, *Nat Prod Rep 28*, 875-896.
37. Bouveng, H., Lindberg, B., and Wickberg, B. (1995) Low-molecular carbohydrates in *algae*, *Acta Chem Scand 9*.
38. Santos, H., Lamosa, P., Faria, T. Q., Borges, N., and Neves, C. (2007) The physiological role, biosynthesis, and mode of action of compatible solutes from (hyper)thermophiles, In *Physiology and Biochemistry of Extremophiles* (Gerday, C., and Glansdorff, N., Eds.), pp 86-103, ASM Press, Washington D.C.

39. Alarico, S., Empadinhas, N., Mingote, A., Simoes, C., Santos, M. S., and da Costa, M. S. (2007) Mannosylglycerate is essential for osmotic adjustment in *Thermus thermophilus* strains HB27 and RQ-1, *Extremophiles 11*, 833-840.
40. Silva, Z., Borges, N., Martins, L. O., Wait, R., da Costa, M. S., and Santos, H. (1999) Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by Rhodothermus marinus and *Rhodothermus obamensis*, *Extremophiles 3*, 163-172.
41. Borges, N., Marugg, J. D., Empadinhas, N., da Costa, M. S., and Santos, H. (2004) Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of *Rhodothermus marinus*, *J Biol Chem 279*, 9892-9898.
42. Empadinhas, N., Albuquerque, L., Henne, A., Santos, H., and da Costa, M. S. (2003) The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate, *Appl Environ Microbiol 69*, 3272-3279.
43. Empadinhas, N., Marugg, J. D., Borges, N., Santos, H., and da Costa, M. S. (2001) Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*. Biochemical and genetic characterization of key enzymes, *J Biol Chem 276*, 43580-43588.
44. Martins, L. O., Empadinhas, N., Marugg, J. D., Miguel, C., Ferreira, C., da Costa, M. S., and Santos, H. (1999) Biosynthesis of mannosylglycerate in the thermophilic bacterium R*hodothermus marinus*. Biochemical and genetic characterization of a mannosylglycerate synthase, *J Biol Chem 274*, 35407-35414.
45. Neves, C., da Costa, M. S., and Santos, H. (2005) Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: osmoadaptation and

thermoadaptation in the order thermococcales, *Appl Environ Microbiol 71*, 8091-8098.

46. Empadinhas, N. (2005) Pathways for the synthesis of mannosylglycerate in prokaryotes: genes, enzymes and evolutionary implications, PhD thesis University of Coimbra, Coimbra, Portugal.
47. Empadinhas, N., Albuquerque, L., Costa, J., Zinder, S. H., Santos, M. A., Santos, H., and da Costa, M. S. (2004) A gene from the mesophilic bacterium *Dehalococcoides ethenogenes* encodes a novel mannosylglycerate synthase, *J Bacteriol 186*, 4075-4084.
48. Flint, J., Taylor, E., Yang, M., Bolam, D. N., Tailford, L. E., Martinez-Fleites, C., Dodson, E. J., Davis, B. G., Gilbert, H. J., and Davies, G. J. (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase, *Nat Struct Mol Biol 12*, 608-614.
49. Lairsonb, L. L., and Withers, S. G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes, *Chem Commun* (Camb, U.K.), 2243-2248.
50. Pesnot, T., Jorgensen, R., Palcic, M. M., and Wagner, G. K. (2010) Structural and mechanistic basis for a new mode of glycosyltransferase inhibition, *Nat Chem Biol 6*, 321-323.
51. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases, *J Mol Biol 328*, 307-317.
52. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases, *Structure 3*, 853-859.
53. Lesk, A. M. (1995) NAD-binding domains of dehydrogenases, *Curr Opin Struct Biol 5*, 775-783.
54. Charnock, S. J., and Davies, G. J. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms, *Biochemistry 38*, 6380-6385.
55. Vrielink, A., Ruger, W., Driessen, H. P., and Freemont, P. S. (1994) Crystal structure of the DNA modifying enzyme beta-

glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose, *EMBO J 13*, 3413-3422.

56. Chiu, C. P., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2004) Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog, *Nat Struct Mol Biol 11*, 163-170.
57. Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases, *Glycobiology 16*, 29R-37R.
58. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annu Rev Biochem 77*, 521-555.
59. Heightman, T. D., and Vasella, A. T. (1999) Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases, *Angew. Chem, Int Ed 38*, 750-770.
60. Zechel, D. L., Reid, S. P., Stoll, D., Nashiru, O., Warren, R. A., and Withers, S. G. (2003) Mechanism, mutagenesis, and chemical rescue of a beta-mannosidase from cellulomonas fimi, *Biochemistry 42*, 7195-7204.
61. Davies, G. J., Sinnott, M. L., and Withers, S. G. (1998) In *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.), pp 119-208, Academic Press, San Diego.
62. Sinnott, M. L. (1990) Catalytic mechanism of enzymic glycosyl transfer, *Chem. Rev.* (Washington, DC, U. S.) *90*, 1171-1202.
63. Pedersen, L. C., Dong, J., Taniguchi, F., Kitagawa, H., Krahn, J. M., Pedersen, L. G., Sugahara, K., and Negishi, M. (2003) Crystal structure of an alpha 1,4-N-acetylhexosaminyltransferase (EXTL2), a member of the exostosin gene family involved in heparan sulfate biosynthesis, *J Biol Chem 278*, 14420-14428.

64. Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases, *Proc Natl Acad Sci U S A 100*, 845-849.

65. Errey, J. C., Lee, S. S., Gibson, R. P., Martinez Fleites, C., Barry, C. S., Jung, P. M., O'Sullivan, A. C., Davis, B. G., and Davies, G. J. (2010) Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors, *Angew Chem, Int Ed Engl 49*, 1234-1237.

66. Lairson, L. L., Chiu, C. P., Ly, H. D., He, S., Wakarchuk, W. W., Strynadka, N. C., and Withers, S. G. (2004) Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue, *J Biol Chem 279*, 28339-28344.

67. Ly, H. D., Lougheed, B., Wakarchuk, W. W., and Withers, S. G. (2002) Mechanistic studies of a retaining alpha-galactosyltransferase from Neisseria meningitidis, *Biochemistry 41*, 5075-5085.

68. Compain, P., and Martin, O. R. (2001) Carbohydrate mimetics-based glycosyltransferase inhibitors, *Bioorg Med Chem 9*, 3077-3092.

69. Waldscheck, B., Streiff, M., Notz, W., Kinzy, W., and Schmidt, R. R. (2001) alpha(1-3)-Galactosyltransferase Inhibition Based on a New Type of Disubstrate Analogue, *Angew Chem, Int Ed Engl 40*, 4007-4011.

70. Sinnott, M. L., and Jencks, W. P. (1980) *J. Am.Chem. Soc.* 10*2*, 20-26.

71. Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R., and Davies, G. J. (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA, *Chem Biol 9*, 1337-1346.

72. Klein, H. W., Im, M. J., and Palm, D. (1986) Mechanism of the phosphorylase reaction. Utilization of D-gluco-hept-1-enitol in the absence of primer, *Eur J Biochem 157*, 107-114.

73. Cleland, W. W., and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer, *Chem Rev 106*, 3252-3278.
74. Anderson, M. A., Shim, H., Raushel, F. M., and Cleland, W. W. (2001) Hydrolysis of phosphotriesters: determination of transition states in parallel reactions by heavy-atom isotope effects, *J Am Chem Soc 123*, 9246-9253.
75. Catrina, I. E., and Hengge, A. C. (2003) Comparisons of phosphorothioate with phosphate transfer reactions for a monoester, diester, and triester: isotope effect studies, *J Am Chem Soc 125*, 7546-7552.
76. Grzyska, P. K., Czyryca, P. G., Purcell, J., and Hengge, A. C. (2003) Transition state differences in hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions, *J Am Chem Soc 125*, 13106-13111.
77. Lad, C., Williams, N. H., and Wolfenden, R. (2003) The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases, *Proc Natl Acad Sci U S A 100*, 5607-5610.
78. Admiraal, S. J., and Herschlag, D. (1995) Mapping the transition state for ATP hydrolysis: implications for enzymatic catalysis, *Chem Biol 2*, 729-739.
79. Cleland, W. W. (1990) Secondary 18O isotope effects as a tool for studying reactions of phosphate mono-, di-, and triesters, *FASEB J 4*, 2899-2905.
80. Cleland, W. W., and Hengge, A. C. (1995) Mechanisms of phosphoryl and acyl transfer, *FASEB J 9*, 1585-1594.
81. Choe, J. Y., Iancu, C. V., Fromm, H. J., and Honzatko, R. B. (2003) Metaphosphate in the active site of fructose-1,6-bisphosphatase, *J Biol Chem 278*, 16015-16020.
82. Baxter, N. J., Hounslow, A. M., Bowler, M. W., Williams, N. H., Blackburn, G. M., and Waltho, J. P. (2009) MgF(3)(-) and alpha-

galactose 1-phosphate in the active site of beta-phosphoglucomutase form a transition state analogue of phosphoryl transfer, *J Am Chem Soc 131*, 16334-16335.

83. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction, *Science 299*, 2067-2071.
84. Williams, N. H. (2004) Models for biological phosphoryl transfer, *Biochim Biophys Acta 1697*, 279-287.
85. Aqvist, J., Kolmodin, K., Florian, J., and Warshel, A. (1999) Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data?, *Chem Biol 6*, R71-80.
86. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold, *Trends Biochem Sci 29*, 495-503.
87. Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liau, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate, *Nature 462*, 739-744.
88. Reitman, Z. J., and Yan, H. (2010) Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism, *J Natl Cancer Inst 102*, 932-941.
89. Yates, S. P., Edwards, T. E., Bryan, C. M., Stein, A. J., Van Voorhis, W. C., Myler, P. J., Stewart, L. J., Zheng, J., and Jia, Z. (2011) Structural basis of the substrate specificity of bifunctional isocitrate dehydrogenase kinase/phosphatase, *Biochemistry 50*, 8103-8106.
90. Cozzone, A. J., and El-Mansi, M. (2005) Control of Isocitrate Dehydrogenase Catalytic Activity by Protein Phosphorylation in *Escherichia coli*, *J Mol Microbiol Biotechnol 9*, 132-146.

91. Holms, H. (1996) Flux analysis and control of the central metabolic pathways in *Escherichia coli*, *FEMS Microbiol Rev 19*, 85-116.

92. Adams, M. J., Ford, G. C., Koekoek, R., Lentz, P. J., McPherson, A., Jr., Rossmann, M. G., Smiley, I. E., Schevitz, R. W., and Wonacott, A. J. (1970) Structure of lactate dehydrogenase at 2.8 A resolution, *Nature 227*, 1098-1103.

93. Hurley, J. H., Thorsness, P. E., Ramalingam, V., Helmers, N. H., Koshland, D. E., Jr., and Stroud, R. M. (1989) Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase, *Proc Natl Acad Sci U S A 86*, 8635-8639.

94. Hurley, J. H., Dean, A. M., Koshland, D. E., Jr., and Stroud, R. M. (1991) Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and $NADP^+$ complexes, *Biochemistry 30*, 8671-8678.

95. Hurley, J. H., and Dean, A. M. (1994) Structure of 3-isopropylmalate dehydrogenase in complex with NAD^+: ligand-induced loop closing and mechanism for cofactor specificity, *Structure 2*, 1007-1016.

96. Garnak, M., and Reeves, H. C. (1979) Phosphorylation of Isocitrate dehydrogenase of *Escherichia coli*, *Science 203*, 1111-1112.

97. Hurleyb, J. H., Dean, A. M., Sohl, J. L., Koshland, D. E., Jr., and Stroud, R. M. (1990) Regulation of an enzyme by phosphorylation at the active site, *Science 249*, 1012-1016.

98. Hurleya, J. H., Dean, A. M., Thorsness, P. E., Koshland, D. E., Jr., and Stroud, R. M. (1990) Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme, *J Biol Chem 265*, 3599-3602.

99. Zheng, J., and Jia, Z. (2010) Structure of the bifunctional isocitrate dehydrogenase kinase/phosphatase, *Nature 465*, 961-965.

100. Xu, X., Zhao, J., Xu, Z., Peng, B., Huang, Q., Arnold, E., and Ding, J. (2004) Structures of Human Cytosolic NADP-dependent Isocitrate

Dehydrogenase Reveal a Novel Self-regulatory Mechanism of Activity, *J Biol Chem 279*, 33946-33957.

101. Taylor, A. B., Hu, G., Hart, P. J., and McAlister-Henn, L. (2008) Allosteric motions in structures of yeast NAD^+-specific isocitrate dehydrogenase, *J Biol Chem 283*, 10872-10880.
102. Doyle, S. A., Beernink, P. T., and Koshland, D. E., Jr. (2001) Structural basis for a change in substrate specificity: crystal structure of S113E isocitrate dehydrogenase in a complex with isopropylmalate, Mg^{2+}, and NADP, *Biochemistry 40*, 4234-4241.
103. Finer-Moore, J., Tsutakawa, S. E., Cherbavaz, D. B., LaPorte, D. C., Koshland, D. E., and Stroud, R. M. (1997) Access to Phosphorylation in Isocitrate Dehydrogenase May Occur by Domain Shifting, *Biochemistry 36*, 13890-13896.
104. Karlstrom, M., Stokke, R., Steen, I. H., Birkeland, N. K., and Ladenstein, R. (2005) Isocitrate dehydrogenase from the hyperthermophile *Aeropyrum pernix*: X-ray structure analysis of a ternary enzyme-substrate complex and thermal stability, *J Mol Biol 345*, 559-577.
105. Peng, Y., Zhong, C., Huang, W., and Ding, J. (2008) Structural studies of *Saccharomyces cerevesiae* mitochondrial NADP-dependent isocitrate dehydrogenase in different enzymatic states reveal substantial conformational changes during the catalytic reaction, *Prot Sci 17*, 1542-1554.

CHAPTER 2

Structural analysis of *Thermus thermophilus HB27* mannosyl-3-phosphoglycerate synthase*

Evidence for a second catalytic metal ion and new insight into the retaining mechanism of glycosyltransferases.

Susana Gonçalves, Nuno Borges, Ana M. Esteves, Bruno L. Victor, Cláudio M. Soares, Helena Santos, and Pedro M. Matias†

From the Instituto de Tecnologia Química e Biologica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

* This chapter is adapted from two publications: [Goncalves, S., Borges,N., Santos, H., Matias, P.M. (2009). Crystallization and preliminary X-ray analysis of mannosyl-3-phosphoglycerate synthase from *Thermus thermophilus* HB27. *Acta Cryst F* 65, 1014-1017], and [Goncalves, S., Borges, N., Esteves, A.M., Victor, B.L., Soares, C.M., Santos, H., Matias, P.M (2010). Structural analysis of *Thermus thermophilus* HB27 mannosyl-3-phosphoglycerate synthase provides evidence for a second catalytic metal ion and new insight into the retaining mechanism of glycosyltransferases. *J Biol Chem* 285, 17857-17868].

† Author´s contributions:

Susana Gonçalves (S.G.) established the expression, purification of mannosyl-3-phosphoglycerate synthase from *Thermus thermophilus* HB27, under the supervision of Dr. Nuno Borges (N.B.) and Dr. Helena Santos (H.S.). Mutagenesis and kinetic assays were also designed and carried out by S.G., under the supervision of N.B. and H.S. The production and purification mannosyl-3-phosphoglycerate phosphatase from *Rhodothermus marinus*, and differential scanning calorimetry assays were conducted by Ana M. Esteves, under the supervision of N.B. and H.S. The crystallization, X-ray data collection, structural determination and refinement were carried out by S.G., under the supervision of Dr. Pedro M. Matias. The molecular docking calculations were carried out by Dr. Bruno Victor and Dr. Cláudio M. Soares.

2.1 ABSTRACT

Mannosyl-3-phosphoglycerate synthase (MpgS) is a key enzyme in the biosynthesis of α-mannosylglycerate (MG).

The three-dimensional structure of MpgS from *Thermus thermophilus* HB27 in its binary complex form, with GDP-α-D-mannose and Mg^{2+}, was solved to 2.9 Å resolution by Multiple Wavelength Anomalous Dispersion (MAD) at the Zinc *K*-absorption edge. Structural analysis confirmed this as falling into the clan of the retaining enzymes with GT-A fold (www.cazy.org). Two flexible loops involved in key interactions for the productive binding of substrates were identified by comparing the apo-form with the metal-substrate complex [Mg^{2+}:GDP-α-D-mannose]. A second metal binding site was found about 6 Å away from the mannose moiety. Kinetic and mutagenesis studies provided evidence that this second metal site is indispensable for catalysis. Additionally, Asp167 of the conserved D-X-D motif was found within van der Waals contact distance of the anomeric C1' cabon atom in the mannopyranose ring, and potentially acts as the catalytic nucleophile, either in the formation of a glycosyl-enzyme intermediate according to the double-displacement S_N2 reaction mechanism, or in the stabilization of the oxocarbenium ion-like intermediate in light of the $D_N{*}A_{Nss}$ (S_Ni-like) reaction mechanism (Fig1.4). A survey of enzymes with the GT-A fold and with a bound sugar-donor, or a bound analogue, was used to identify the orientation of the scissile glycosidic bond of the sugar donor with respect to the D/E-X-D motif. The glycosidic bond is oriented towards the motif in the retaining GTs and away from it in the inverting GTs. This feature, which explains the stereochemistry of the reactions, provides a structural signature that will assist efforts in classifying the inverting and retaining GTs.

2.2 INTRODUCTION

Mannosyl-3-phosphoglycerate synthase (MpgS; EC 2.4.1.217) is involved in the synthesis of the glycoside α-mannosylglycerate (MG) via a two-step pathway, and it has been classified within the retaining glycosyltransferases GT55 family (www.cazy.org).

Glycosyltransferases (GTs) catalyse the synthesis of an immense diversity of oligosaccharides, polysaccharides, and glycoconjugates, which are key-players in vital processes, such as signaling, storage, stress response and cell structure (*1*).

Despite the success of the techniques applied to the retaining glycoside hydrolases (GHs), in trapping the intermediary catalytic state and in identifying the catalytic nucleophile (*2*), similar approaches have thus far been unsuccessful for the retaining GTs. In contrast with the proposed mechanisms for inverting GTs that parallel a single direct displacement S_N2-like reaction as established for the inverting GHs (*3*), the retaining mechanism of glycosyl transfer is less obvious, as there is less stringency for a common scaffold at the β-face of the sugar donor, in order to fully support a double displacement S_N2-like reaction, similarly to the retaining GHs (*4*). Instead, a front-face S_Ni-like mechanism, was proposed for the retaining GTs, and later designated as $D_N{}^*A_{Nss}$ for GTs according to the IUPAC nomenclature. This topic has been extensively researched and a recent comprehensive review is available (*5*). The structural studies of the retaining α-1,4-galactosyl transferase (LgtC) in complex with donor and sugar acceptor analogues (*6*) are one of the pivotal works in the scrutiny of the retaining GT mechanism, an argument in favor of the S_Ni-like mechanism being provided by the position of Gln198 (*6*). Alternatively, studies made on the bovine α-1,3-Galactosyltransferase (α-GalT) point towards formation of a glycosyl-enzyme intermediate, with Asp317 as the likeliest catalytic nucleophile (*5, 7*). Surprisingly, this enzyme revealed inverting activity when transferring galactose to water or azide (*7*). Evolutionary studies argue that the retaining GT-A enzymes have evolved from an inverting

ancestor (*5, 8*) and recent work has addressed the elucidation of representative structural scaffolds for both inverting and retaining GTs (*4, 5, 9*). Currently, the double-displacement S_N2-like cannot be totally disregarded in favor of the front-face S_Ni-like mechanism.

In this work, we describe the three-dimensional structure of *T. thermophilus* HB 27 MpgS in three crystal forms: its apo-form and binary complexes with GDP-Man:Mg^{2+} of the wild-type protein and its H309A mutant. The initial structural results prompted additional studies which revealed the existence of a co-catalytic metal centre in *T. thermophilus* HB27 MpgS. The protein residues coordinating this metal belong to a conserved sequence motif in all MpgS enzymes of known sequence. In addition, Asp167 in the DXD motif has been identified as a possible catalytic nucleophile in retaining GTs.

2.3 MATERIALS AND METHODS

Expression and purification of the MpgS from T. thermophilus *HB27*

A plasmid containing the *mpgS* gene (Accession; YP_004563) cloned in pKK223-3, was kindly provided by M. S. da Costa, Coimbra, Portugal. Transformation, expression and purification of MgpS was performed as previously described (*10*) with some modifications. *E. coli* BL21-Rosetta strains (Novagen), bearing the construction, were grown at 37ºC in YT medium supplemented with ampicillin (100 μg/*m*L) and cloramphenicol (40 μg/*m*L) to an OD_{600} of 0.6-0.7, and induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 6 hours. The cells were harvested by centrifugation (7,000 × g, 15 min, 4ºC) and suspended in Tris-HCl buffer (20 *m*M, pH 7.6) containing 5 *m*M $MgCl_2$ and a cocktail of protease inhibitors (Roche). Cells were disrupted in a French press and DNA was digested with benzonase (2 U/*m*L, Novagen). Cell debris was removed by centrifugation (18,000 × g, 40 min, 4ºC), and the cell extract heated for 10 min at 80ºC to precipitate thermo-labile proteins. After centrifugation (25,000 x g, 45 min, 4ºC), the supernatant was applied into a 70 *m*L Q-sepharose column (GE Healthcare) equilibrated with buffer A (20 *m*M Tris-HCl, pH 7.6). Elution was carried out with a linear gradient from buffer A to buffer B (20 *m*M Tris-HCl pH 7.6, 1 M NaCl). Fractions containing MpgS were eluted between 300 and 350 *m*M NaCl. The fractions were pooled and dialyzed against buffer A before loading into a 6 mL Resource Q column (GE Healthcare). MpgS eluted between 300 and 350 m*M* NaCl of buffer B, and the purest fractions were pooled and dialyzed against buffer A before loading into a 1 mL Mono Q column (GE Healthcare). The fractions eluted at 350 *m*M NaCl were pooled and loaded into a 24 *m*L Superdex 75 10/300 GL (GE Healthcare) equilibrated with buffer A containing 350 *m*M NaCl. In all chromatographic steps, the MpgS activity was detected by visualizing the formation of MG, derived from MPG after treatment with alkaline phosphatase, by Thin Layer Chromatography (*10*). The purified protein was judged pure by SDS-PAGE (12%), showing a single band with an apparent molecular mass of

43.6 kDa (Figure 2.1A). Protein concentration was determined by the Bradford assay (*11*). The molecular mass of MpgS was estimated by gel filtration (Figure 2.1B) using a 2.4 *m*L Superdex 200 3.2/30 PC column (GE Healthcare) equilibrated with 10 *m*M potassium phosphate buffer (pH 7.2) containing 350 *m*M NaCl. Ribonuclease (13.7 kDa), chymotrypsinogen (25 kDa), ovalbumin (43 kDa), albumin (66 kDa), aldolase (158 kDa), and ferritin (440 kDa) were used as standards (GE Healthcare). Blue dextran 2000 (GE Healthcare) was used to determine the void volume of the column. MpgS eluted at the same volume (1.56 *m*L) as albumin, suggesting that a homodimeric structure is the prevalent oligomeric state under the running conditions.

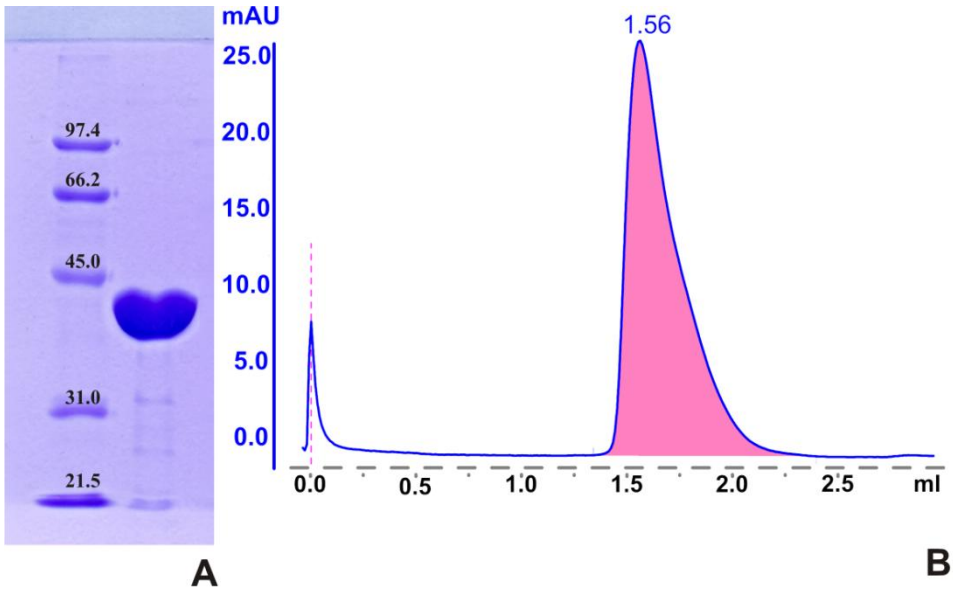


Figure 2.1. A) - SDS-PAGE 12% with pure recombinant MpgS from *Thermus thermophilus* HB27. Lane 1: low Range Molecular Weight Marker (BioRad) with protein markers from the highest to the lowest mass; phosphorylase b (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), lysozyme (14.4 kDa). Lane 2: MpgS monomer migrating accordingly to its molecular weight (43.5 kDa); **B) -** Elution profile of MpgS loaded into an analytical 2.4 *m*L Superdex 200 3.2/30 PC column. The elution volume suggests that the protein is most likely in the dimeric state. mAU stands for mili units of absorption.

MpgS crystallization and cryo-conditions

Preliminary crystallization trials were first done at the *nano*scale with the commercially available kits Crystal Magic I and II (Biogenova), using the Cartesian Crystallization Robot Dispensing System (Genomics Solutions) with round-bottom Greiner 96-well CrystalQuick^{TM} plates (Greiner Bio-One). Three crystallization drops per condition screened were prepared, using 100 *n*L reservoir solution and 100 *n*L protein solution at three different concentrations: 5 *m*g/*m*L, 10 *m*g/*m*L and 17 *m*g/*m*L. The drops were equilibrated against 100 μL of reservoir solution. Crystals with flat faces and sharp edges were often found in the presence of several kinds of PEGs (polyethylene glycol), in a buffering system with pH between 7 and 9 in all three concentrations, but with the highest concentration yielding larger crystals. Although nicely shaped, these were hard to handle and very little diffraction was observed. When studying the effect of divalent metals for crystallization, a new crystal form was found, developing strictly when adding $ZnCl_2$ within a narrow concentration interval of 400 to 600 μM in the crystallization solution of Crystal Magic II condition 5 consisting of 0.2 M Magnesium Acetate tetrahydrate (MgAcetate), 0.1 M Sodium Cacodylate (NaCacodylate) pH 6.5, 30% 2-methyl-2,4-pentanediol (MPD). This result was reproducible and was used during crystal growth by scaling-up. Native protein was concentrated up to 17 *m*g/*m*L and 2 μL drops (in a 1:1 ratio of protein to reservoir solution) were set up at 20ºC in 24 well crystallization plates using the

hanging-drop vapour diffusion technique. The drops were equilibrated against 500 μL of reservoir solution, composed of 0.2 M MgAcetate, 0.1 M NaCacodylate pH 6.5, 30%-35% MPD and 600 μM $ZnCl_2$ as additive. Prismatic crystals, as illustrated in Figure 2.2, developed within 5 days in both cases, with dimensions varying between 100 and 400 μm in the longest axis. Crystals of the binary complexes; wtMpgS:GDP-Man:Mg^{2+} and H309A MpgS:GDP-Man:Mg^{2+} , were obtained by co-crystallization by adding 5 *m*M $MgCl_2$ and 5 *m*M of GDP-Man to the protein solution prior the setting up of the crystallization drops.

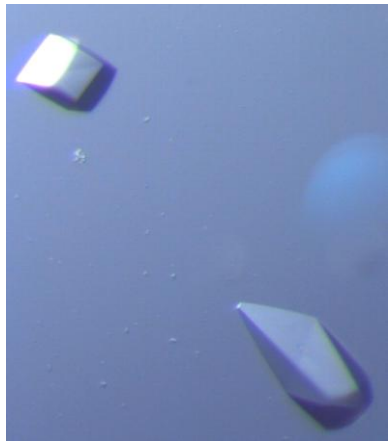


Figure 2.2 - Prismatic crystals of *T. thermophilus* HB27 MpgS

Crystal cryo-protection

For data collection, the crystals were initially cryo-protected prior to flash-cooling in LN_2 by transferring them directly into reservoir solution without Zn^{2+} and supplemented with 12.5 % glycerol. The highest resolution observed was about 2.9 Å. The diffraction limit was independent of both crystal age and cryo-treatment, as room temperature in-house data collection (Bruker AXS Proteum diffractometer and μStar rotating anode generator, Montel mirrors, CuKα radiation) showed no improvement of diffraction resolution. However, the cryo-treatment affected the internal crystal order, as revealed by spot mosaic spread and by the higher background level of the diffraction pattern. After several tests, an improved flash-cooling procedure was used where the

crystallization buffer was replaced by sodium citrate, and consisted of two steps: the crystals were first transferred into a solution containing 0.2 M MgAcetate, 0.2 M NaCitrate pH 6.5, 32.5 % MPD, and left to equilibrate for about 3 minutes; next, they were dipped in a cryo-solution with the same composition as in the first step and supplemented with 5% glycerol, and immediately flash-cooled in LN_2. Spot quality improved significantly, a lower background was observed in the diffraction pattern, but no increment in the resolution limit was achieved.

Data collection and processing

Crystals of the binary complexes wild-type (wt) MpgS:GDP-Man:Mg^{2+} and mutant H309A MpgS:GDP-Man:Mg^{2+} were obtained by co-crystallization. A 3-wavelength Multiple Wavelength Anomalous Dispersion (MAD) data set to 3.0 Å resolution was measured at the Zinc absorption edge K from a flash cooled crystal of wtMpgS:GDP-Man:Mg^{2+} at 100 K on ESRF beamline ID-29. A dataset from another wtMpgS:GDP-Man:Mg^{2+} crystal was measured to 2.8 Å at 100 K on ESRF beamline ID23-1. A dataset from an apo-MpgS crystal was measured to 3.0 Å at 100 K on ESRF beamline ID14-4. A dataset from a H309A MpgS:GDP-Man:Mg^{2+} crystal was measured to 3.0 Å at 100 K on DLS beamline I04. Diffraction images were processed with the XDS Program Package (*12*). The diffraction pattern was seen to suffer from anisotropy, which resulted in the unusually large values of R_{merge} and R_{meas}, despite a reasonable value for $\langle I/\sigma(I)\rangle$ of around 1.3 to 2 in the last shell, with the exception of the remote dataset which had a value for $\langle I/\sigma(I)\rangle$ of 0.6. A summary of the data collection statistics is presented in Table 2.1.

Structure determination

Using the HKL2MAP (*13*) graphical user interface, the 3-wavelength MAD dataset was scaled and analysed with SHELXC, the Zn^{2+} heavy atom substructure was determined with SHELXD (*14*) and the phase problem solved with SHELXE (*15*). The best solution from SHELXD in 100 trials gave 5 Zn^{2+}

sites with a correlation coefficient of 38.3%, two of them with apparently about half the occupancy of the top three. The SHELXE calculations gave a clear discrimination between the correct and the inverted substructure solutions, and resolved the space-group ambiguity in favor of $P4_12_12$. The phases derived from the MAD data were further improved, using the maximum-likelihood heavy-atom parameter refinement in SHARP (*16*), with resolution extension using the ID23-1 2.8 Å dataset, and followed by density modification with SOLOMON (*17*). An optimizing density modification procedure using SOLOMON suggested a solvent content of 62%, which was used in all subsequent density modification calculations. Centroid SHARP phases were further improved by density modification with DM (*18*) using the CCP4i Graphical User Interface (*19, 20*), and with non-crystallographic symmetry (NCS) averaging. No dimeric arrangement for the two MpgS molecules in the asymmetric unit could be initially perceived from the spatial arrangement of the five Zn^{2+} positions. Therefore, a preliminary NCS operator relating the two monomers was calculated with LSQKAB (*21*) by superposing two triangles derived from the five Zn^{2+} positions. At this point, a random 5% sample of the reflection data was flagged for R-free calculations (*22*) during model building and refinement. Initial model building and sequence docking was done automatically with Buccaneer (*23*), and a preliminary model was completed with Coot (*24*). At this stage, the dimeric arrangement of MpgS in the asymmetric unit could be clearly discerned, and a better NCS operator was derived, corresponding to a two-fold NCS axis at the dimer interface. In parallel, and after inspection of the anomalous residual maps, a second SHARP/SOLOMON run was carried out based on the previous one, with anisotropic thermal motion B-factor correction for the Zn^{2+} sites, which gave better phases as judged by the figure of merit. Phases from this procedure were again improved in DM/CCP4i using the optimized 2-fold NCS operator. Phasing and phase refinement statistics, with and without anisotropic B-factor correction, are listed in Table 2.2. Using the new phases from DM, an automated model building, sequence docking and refinement procedure was

carried out with Buccaneer/REFMAC (*18, 25*). Finally, an electron density map, calculated using the 2.8 Å dataset and the improved DM phases was used for completing the model with Coot.

The structures of apo-MpgS and H309A MpgS:GDP-Man:Mg^{2+} were determined by molecular replacement with PHASER (*26*) using the refined protein chain coordinates of wtMpgS:GDP-Man:Mg^{2+}.

Crystallographic refinement

The three structures were refined using a Maximum-Likelihood refinement procedure with REFMAC (*25*), as implemented through the CCP4i graphics user interface (*19*). Non-crystallographic symmetry restraints between the two independent molecules in the asymmetric unit were applied throughout the refinement, and a Translation-Libration-Screw (TLS) rigid body refinement (*20*) was carried out prior to restrained refinement of atomic positions and isotropic thermal motion parameters. One rigid body was defined for each of the two independent molecules in the asymmetric unit including its bound ligand GDP-Man. The electron density for three flexible loop regions designated as Loop1, Loop2 and Loop3 respectively in each monomer (see "Results") and corresponding to the residue ranges Tyr203 - Val214, His311 - Gly316 and Gly352 - Glu359, was not interpretable in some cases for Loops 1 and 3 as shown in Table 2.3 and therefore not included in the model. The same reflection indices retrieved from the wtMpgS:GDP-Man:Mg^{2+} dataset were used for subsequent R-free calculations during the structural refinements of the apoMpgS and H309A MpgS structures. The wtMpgS:GDP-Man:Mg^{2+} was refined using the ID23-1 2.8 Å dataset to final values of R-factor and R-free of 0.176 and 0.234, respectively. The apo-MpgS:GDP-Man:Mg^{2+} structure was refined using the ID14-4 3.0 Å dataset to final values of R-factor and R-free of 0.198 and 0.266. The H309A MpgS:GDP-Man:Mg^{2+} structure was refined using the I04 3.0 Å dataset to final values of R-factor and R-free of 0.174 and 0.232. During the refinement process, the models were periodically inspected and corrected in a 3D Graphics workstation with Coot (*24*) against

SigmaA-weighted $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps, which were also used to include water molecules in the refinement. The final refinement statistics are shown in Table 2.3. Model stereochemical quality was assessed with PROCHECK (*27*). The secondary structure was analysed with PROMOTIF v 3.0 (*28*) and PROCHECK (*27*).

Coordinates and structure factors were deposited in the RCSB Protein Data Bank (*29*) with accession numbers 2wvk and r2wvksf for apo-MpgS, 2wvl and r2wvlsf for wtMpgS:GDP-Man:Mg^{2+}, 2wvm and r2wvmsf for H309A MpgS:GDP-Man:Mg^{2+}.

Mutagenesis studies

The production of the E251A and H309A MpgS mutant proteins was done by *in vitro* site-directed mutagenesis, using the QuickChange Site-Directed Mutagenesis kit (Stratagene), and following the manufacturer's instructions. The *mpgS* gene cloned in the pKK223-3 vector was used as template for plasmid amplification by Polymerase Chain Reaction (PCR). The mutagenic primers used were: E251A-sense 5`-CG AAC GCC GGG GCG CAC GCC ATG AG-3` and E251A-antisense 5´-CT CAT GGC GTG CGC CCC GGC GTT CG-3´, for the E251A substitution; and the primers H309A-sense 5'-G ACC CGG AAC CCC GCC CTC CAC GAG AAC-3' and H309A-antisense 5´-GTT CTC GTG GAG GGC GGG GTT CCG GGT C-3´, for the H309A substitution. Confirmation of the mutations was done by DNA sequencing (AGOWA Genomics). Expression and purification followed the same procedure as described for the native MpgS and the final products were confirmed by N-terminal sequencing (Analytical Services Unit, ITQB, Portugal). The stability of both mutant proteins was assessed by calorimetry. The observed melting temperatures were similar to that measured for wtMpgS (T_m = 81ºC): T_m = 82ºC for the E251A mutant protein and T_m= 81ºC for the H309A mutant protein.

Production and purification of the Rhodothermus marinus *MpgP for kinetic assays*

The gene from *R. marinus* DSM4252 coding for MpgP, previously cloned in a pKK223-3 vector (*30*), was produced to be further used in the substrate specific dephosphorylation of MPG, during the enzymatic activity assays for the native MpgS and its H309A and E251A mutant proteins. Expression and purification was carried out as described for the *T. thermophilus* HB27 MpgS, but instead of the last anionic exchange chromatographic step (MonoQ), a cationic exchange step was performed (MonoS, Amersham Biosciences) using 20 *m*M MES-NaOH pH 6.5 as the buffer system for protein elution through a salt linear gradient. The purest fractions eluted at about 200 *m*M NaCl and were used for activity assays.

Activity assays

Prior to the activity assays, all protein samples were dialyzed against 20 *m*M Tris-HCl pH 7.6, 25 *m*M NaCl, 10 *m*M EDTA, to ensure maximal removal of any contaminant metal present in solution. A second dialysis step against 20 *m*M Tris-HCl, pH 7.6, was subsequently done for EDTA removal.

Activity detection of the mutants was done by Thin Layer Chromatography (TLC) using wtMpgS as positive control. The reaction mixtures contained 2.5 *m*M each of GDP-Man and 3-PG in 20 *m*M Tris-HCl pH 7.6, 20 *m*M $MgCl_2$ (buffer B1) and 10 μg of purified protein, and were incubated at 70ºC for 1 hour. Formation of the MPG product was followed by its subsequent dephosphorilation with 4 units of bovine intestinal alkaline phosphatase (SIGMA), yielding mannosylglycerate, which was further detected with α-naphtol as previously described by Empadinhas *et al*. (*10*). A significant decrease in color intensity in comparison with the positive control was observed for the H309A mutant protein, while the E251A mutant protein produced a very faint pink color (data not shown).

Determination of the specific activities (μmol.min^{-1}.mg^{-1}) for the wild type enzyme and the mutant proteins was based on the release of inorganic

phosphate from MPG, using a modified protocol from that previously described by Empadinhas and co-workers (*31*) wtMpgS was used for establishing optimized reaction conditions at the optimal temperature of 75ºC, and as positive control in the subsequent assays. Maximum enzyme activity was obtained by using 0.5 μg of pure wtMpgS with an excess (5 *m*M) of GDP-Man and 3-PG in reaction buffer B1 (total volume 50*u*L). An excess (5 μg) of substrate-specific MpgP was added to ensure the complete dephosphorilation of MPG. The quantification of inorganic phosphate release was based on the Ames method (*32*). For wtMpgS, the effect of Zn^{2+} on the specific activity as well as its competing effect on Mg^{2+} were studied by performing the same kinetic assays, using 300 μM of $ZnCl_2$ in Tris-HCl buffer (buffer B2) for the former, and 300 μM of $ZnCl_2$ and 20 *m*M of $MgCl_2$ in Tris-HCl (buffer B3) for the latter. The Zn^{2+} effect in the activity of MpgP was negligible, as assessed by dephosphorylation assays with addition of 300 μM $ZnCl_2$ to the MpgP reaction buffer B4 (25 *m*M Tris-HCl, 10 *m*M $MgCl_2$). Determination of specific activities for the MpgS mutant proteins was done as for the wtMpgS, using instead 5 μg of pure protein to ensure a measurable inorganic phosphate release from MPG. All assays using reaction buffer system B1 were done in triplicate and assays with buffers B2 and B3 were done in duplicate. All solutions were prepared with ultrapure deionized water (Milli-Q, Millipore Corporation).

Molecular modelling

Partial charges were calculated for the catalytic centre of MpgS represented in Fig. 2.9, based on electrostatic potentials calculated using quantum chemistry methods using Gaussian03 (*33*) and RESP fitting (*34*). Initial coordinates for the centre were taken from the X-ray structure of the binary complex. Protons were initially placed by geometry and their conformation was optimized while heteroatoms were kept rigid. This optimization was done using the semiempirical method PM3 (*35*). This was the starting structure for single point density functional B3LYP calculations using the 6-31G(d) basis set. Electrostatic potentials in space were calculated using

this method and the RESP fitting was applied to heteroatoms and polar hydrogens.

The docking simulations of 3-PG into the active site of the MpgS binary complex structure were carried out with Autodock 4.2 (*36*). This software allows the incorporation of flexibility on both the substrate and protein. However, in our simulations we have only applied flexibility into the substrate. The calculations were setup with AutodockTools 1.5.4 (*36*). This program provides an interactive method for defining not only the torsion tree, but also the partial atomic charges for a given ligand and protein. Gasteiger charges were assigned to the 3-PG atoms and ligand flexibility was introduced, with a total number of 6 active torsions. Kollman charges were assigned to all protein atoms, except those in the catalytic centre as described above. The docking experiments were performed with the Lamarkian genetic algorithm (LGA). Using a grid with 52x52x44 points, with a grid spacing of 3.75 Å and centered in the catalytic pocket, we have generated 200 different complexes using the default parameters from Autodock 3 (*36*). A maximum of 25 million energy evaluations were applied for each experiment. The results were clustered using a 2.0 Å tolerance. A second 3-PG was docked in the catalytic pocket while keeping the best docking solution of the first 3-PG, to search for additional docking sites closer to the sugar donor. The charge set used for both protein and ligand in this simulation, as well as the docking parameters, were the same as in the first simulation. The results are shown in Fig. 2.9 and Fig. 2.10.

2.4 RESULTS

The structure of MpgS

Both MpgS:GDP-Man:Mg^{2+} binary complexes from *T. thermophilus* HB27 (wild-type and H309A mutant), as well as the apo-MpgS, crystallized in tetragonal space group $P4_12_12$, with a homodimer in the asymmetric unit. The dimeric assembly, represented in Fig. 2.3, is consistent with the prediction by size exclusion chromatography, and likely corresponds to the biological unit in solution. Each monomer displays an overall globular α/β topology with an extended central β-sheet composed of 9 β-strands in the order β2, β3, β12, β10, β11, β9, β6, β7 and β8, with β2 and β11 running in anti-parallel direction, and its connectivity (*37*) can be described as 0 1 7X 1X 1X -3X -2X 1 -2, considering strand β2 as the origin. The architecture of each monomer (Fig. 2.4), suggests the presence of 3 distinct structural regions. The N-terminal domain comprises two of these regions: a mixed 5-stranded β-region (residues 1 to 43) composed by strands β1 to β5, and a Rossmann-like $\alpha/\beta/\alpha$ sandwich region (residues 55 to 172), both linked to a central helix α1 (residues 45 to 52), which is oriented obliquely to plane of the central β-sheet. The C-terminal domain (residues 173 to 391) begins immediately after the DXD motif (Asp167-Ala168-Asp169), typically present among the metal-dependent GT-A enzymes (*5*), includes the second half of the β-sheet, as well as a 4-helical parallel bundle composed by helices α7, α9, α11, and α12, which encloses part of the catalytic pocket, and ends with a long C-terminal tail that folds back onto the Rossmann-like N-terminal region. This is a representative structure of the GT55 family, together with the recently available selenomethionine derivative of *Pyrococcus horikoshii* OT3 MpgS (PDB 2ZU7).

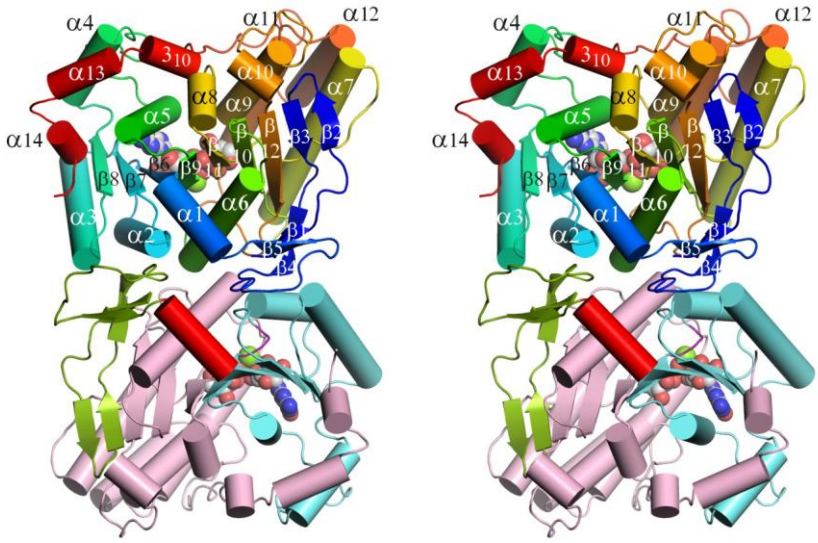


Figure 2.3 *The* structure of MpgS. Stereo cartoon representation of the overall structure and assembly of the MpgS:GDP-Man:Mg^{2+} homodimer. α-helices are represented as cylinders, β-strands as arrows, GDP-Man and Mg^{2+} in space-filling mode. Atom colors are light grey for carbon, blue for nitrogen, red for oxygen, orange for phosphorus and green for Mg^{2+}. The A monomer (bottom) is colored according to the domain regions mentioned in the text: N-terminal β-region (light green), central α-helix (red), β/α/β Rossmann-like region (light blue), DAD motif (purple) and C-terminal α-helical region (light pink)**.** In the B monomer (right) the secondary structure elements are colored from blue (N-terminal) to red (C-terminal). The two monomers are related by a two-fold NCS axis perpendicular to the plane of the figure. Figure prepared with PyMol (http://www.pymol.org).

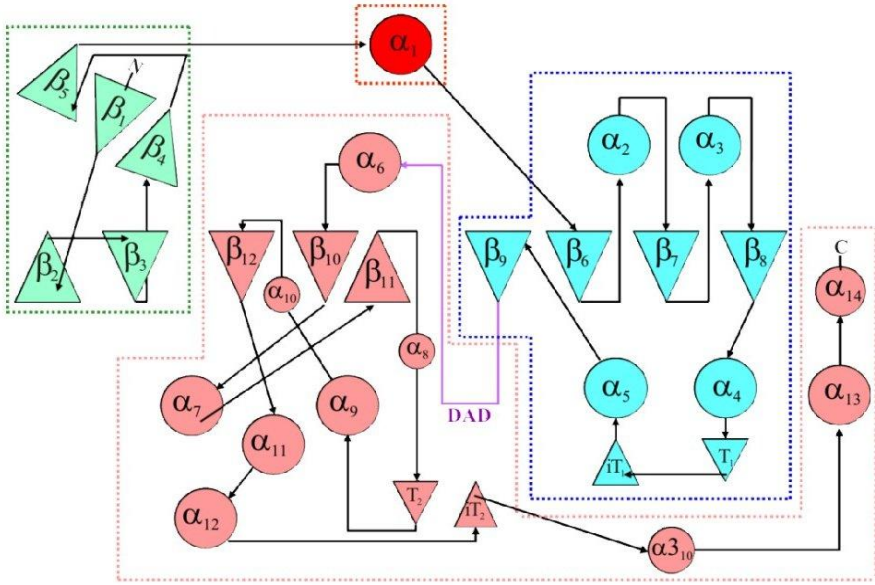


Figure 2.4 Topology diagram of the MpgS monomer. The secondary structure elements are color coded as in Fig. 2.3; the colored boxes represent the domain regions mentioned in the text and shown for one of the MpgS monomers in *A*. α-helices are represented by circles, β-strands by triangles. End-arrows, labeled as T_1 and T_2, correspond to β-bridges and to extended β-strands respectively. Secondary structure elements were assigned based on PROCHECK (*27*) and PROMOTIF (*28*) classification. Topology diagram was based on the TOPS program (*12*).

Structural role of Zn^{2+}

The use of $ZnCl_2$ as an additive was essential to obtain crystals of MpgS. Using the same crystallization conditions as for apo-MpgS, crystals of wtMpgS and its H309A mutant with bound GDP-Man and Mg^{2+} were obtained by co-crystallization and exhibited the same crystal habit and similar cell parameters (Table 2.1). The 5-atom Zn^{2+} substructure determined from the MAD dataset contained three sites with a clearly higher occupation than the other two. Analysis of the refined wtMpgS:GDP-Man:Mg^{2+} crystal structure showed the two minor Zn^{2+} sites to be internal and close to the ligand binding pocket, while the three major sites were located on the surface of the MpgS molecules (Fig. 2.5). One (Zn5) is located at a NCS 2-fold axis (different from that associated with the biological dimer) and is coordinated by Asp292 and His295 residues from each monomer. The other two are 2-fold NCS-related and coordinated by His8 and Glu19 residues from each monomer (Zn1 to monomer A and Zn2 to monomer B) and a water molecule. However, while Zn2 is also coordinated by Asp242 from the same monomer B, Zn1 is additionally coordinated by His105 from monomer B. This coordination scheme leads to stronger intermolecular interactions between adjacent biological dimers which assemble into a tetramer of dimers, and is likely to be the structural foundation behind the success of the MpgS crystallization using $ZnCl_2$ as an additive. This should not be surprising, since Zinc is a particularly prevalent transition metal in protein structures, is frequently involved in protein-protein interaction sites and its affinity for a broad range of ligands from thiolate through imidazolate to carboxylate suits the role of a bridging cation. The redox inactivity of Zn^{2+} also likely contributes to its abundance as an interaction mediator in biological systems (*38*). Supporting this evidence is also the crystallographic structure from *P. horikishii*, where such strategically equivalent coordinating sites are absent and being mostly replaced by neutral and acidic residues with no propensity to establish close contacts via metal-coordination, and allowing it to crystallize without the presence of Zn^{2+} or other metals with a borderline chemical nature .

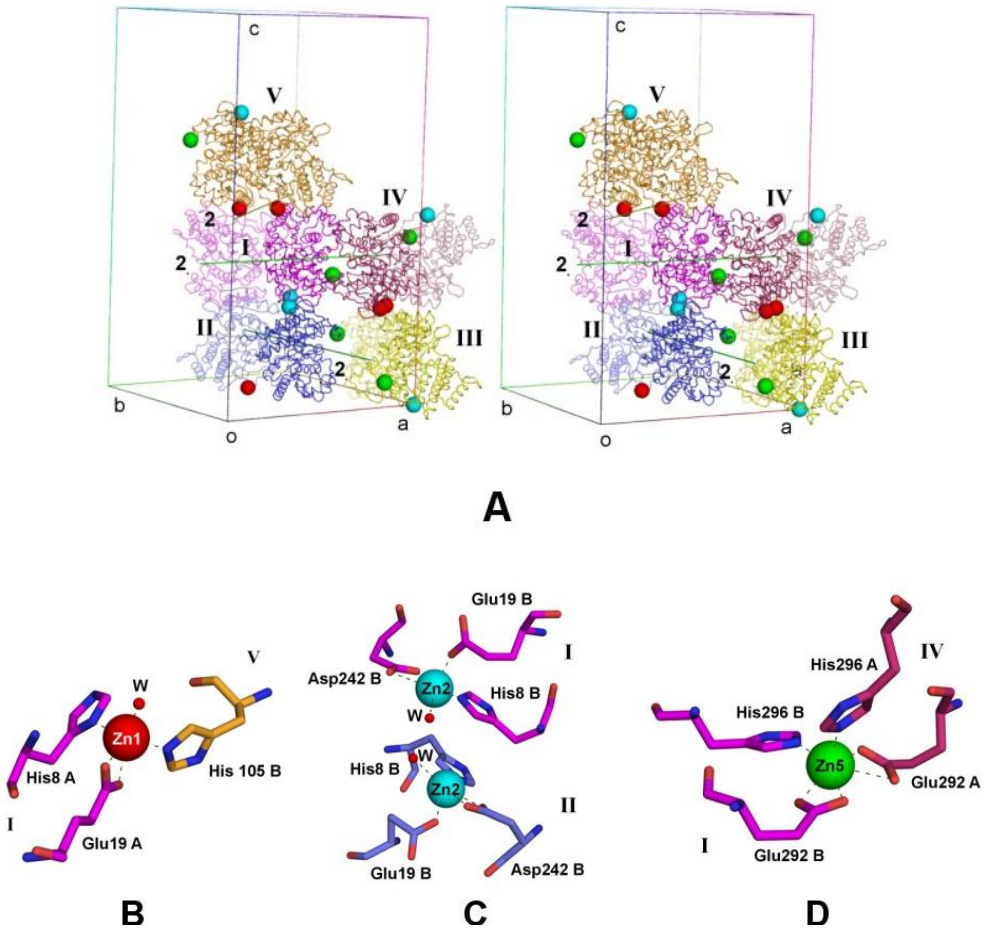


Figure 2.5 The structural role of Zn^{2+}. A) Stereo-view of symmetry-related MpgS homodimers and their coordinated structural Zn^{2+} ions. For clarity, the internal ligands were omitted. Four symetry-related homodimers: I (magenta), II (TV-blue), III (yellow) and IV (raspberry red) define the polymerization unit leading to the spatial arrangement of the crystal structure in space group $P4_12_12$. The 2-fold screw axes parallel to the unit cell axes ***a*** and ***b*** transform homodimers I into IV, and II into III, respectively. Homodimer V (bright orange) is related to homodimer I by a 2-fold symmetry axis parallel to the (001) face diagonal. The biological homodimers are represented as tubes and the Zn^{2+} ions as spheres: Zn1, red; Zn2, cyan and Zn5, green; **B**) Detail view of the Zn1 site: this Zn^{2+} ion is coordinated by His8 ($N^{\delta 1}$) and Glu19 (*syn* $O^{\varepsilon 1}$, *syn* $O^{\varepsilon 2}$) in

monomer A, and by His105 ($N^{\varepsilon 1}$) in a symmetry-related monomer B. An additional water molecule completes the coordination sphere with distorted triangular bipyramidal geometry. **C**) Detail view of the Zn2 site: this Zn^{2+} ion is coordinated by His8 ($N^{\varepsilon 2}$), Glu19 (*anti* $O^{\varepsilon 2}$) and Glu242 (*syn* $O^{\varepsilon 1}$, *syn* $O^{\varepsilon 2}$) in monomer B and by a water molecule (W6), in a distorted square piramidal geometry. A second Zn2 site is adjacent, and related by crystallographic symmetry. **D**) Detail view of the Zn5 site: this site is located on an NCS axis different from that relating the two monomers (A and B) in the biological homodimer. The Zn^{2+} ion is coordinated by Glu292 (*syn* $O^{\delta 1}$, *syn* $O^{\delta 2}$) and His296 ($N^{\varepsilon 2}$) from monomer B and a symmetry-related monomer A. The coordination geometry is distorted octahedral. Residues are shown as sticks and colored according to their respective homodimer as in **A**). Zinc atoms are shown as spheres and colored as in **A**). Water molecules are represented as red spheres. Figure prepared with PyMol (http://www.pymol.org).

Structural homologues of MpgS

Secondary structure comparison against all PDB using the DALI (*39*) search engine, revealed a general structural preservation with many other nucleoside-diphosphosugar (NDP-sugar) transferases, including those involved in the synthesis of lipopolysaccharides, and to some pyrophosphorylases present at the early stages of peptidoglycan biosynthesis. However, and as expected, the highest Z-score (52.9) was observed for its orthologous family member *P. horikoshii* MpgS, followed by members of the GT81 family: *Mycobaterium tuberculosis* glucosyl-3-phosphoglycerate synthase (GpgS) (*40*) (PDB 3E26; Z-score=21.9), the putative GpgS from *Mycobacterium avium paratuberculosis K-10* encoded by MAP2569c (PDB 3CKN; *Z*-score = 21.8) (*41*), and MpgS/GpgS from *Rubrobacter xylanophilus* that is able to use GDP-Man as well as GDP-glucose as substrates (PDB 3F1Y; *Z*-score=21.2) (*42*). This last enzyme is suggested to be at the evolutionary split from which all MpgSs and GpgSs have evolved (*31*). *Rhodotermus marinus* MgS, representative of the GT78 family and involved in

the alternative single-step pathway for mannosylglycerate synthesis (*4, 30*), is also one of the most closely related enzymes (PDB 2BO8; Z-score=19.0). Indeed, these can all be included in the MgS-like fold family (SCOP v.1.75; c.68.1.18) (*43*), suggesting a common structural core composed of the central β-sheet and defined by strands β12, β10, β11, β9, β6, β7 and β8 (MpgS nomenclature; Fig.2.4) as a reminiscent feature from a common ancestor.

Contrary to the α/β/α conserved subdomain, the α-rich C-terminal region is highly variable, is known to be associated with the sugar-acceptor binding specificity, and is also likely related to the protein solubility (*4*). The helical bundle is linked to the β-sheet core through flexible loops; Loop 1 (Tyr203-Val214) and Loop 2 (His311-Gly316) which are involved in the substrate binding mediated response (*4, 40, 41*). In addition to these two highly mobile regions, a third (Gly352-Glu359) is located at the beginning of the long C-terminal tail.

The two MpgS monomers in the asymmetric unit are related by a 2-fold NCS axis at the dimer interface, which is defined by the tight packing of the N-terminal β-region (residues 22-32) with the flexible loop regions of the α/β/α "Rossmann-like" sandwich, comprising mainly the loops connecting α2 to β7 (residues 70-79) and α3 to β8 (residues 107-109) (Figure 2.3). The dimerization interaction is mostly electrostatic, and its effect is potentiated through the solvent accessible residues of the two interlocking α1 helices (residues 45-50), and also by π-stacking between NCS-related pairs of Trp177 and Phe172 residues. The dimer interaction surfaces were analyzed with PISA (*44*). In these calculations, the contributions from bound GDP-Man and metal ions were excluded. Upon dimer formation, an average of 1670 $Å^2$ are buried in each monomer, representing about 10 % of its total solvent accessible surface area. In addition, an estimated total of 32 hydrogen bonds and 36 salt bridges are formed across the dimer interface. These interactions very likely contribute to its stabilization and play an essential role in the assembly of the biological unit of *T. thermophilus* HB27 MpgS, similarly to *P. horikoshii* MpgS. On the other hand, the biological unit proposed for GpgS and MAP2569c (*40, 41*),

involves a two-fold crystallographic axis at the C-terminal sub-domain, which is also seen in the biological dimer of *R. xylanophilus* MpgS/GpgS.

The catalytic pocket

In the crystal structure of GDP-Man:Mg^{2+} binary complexes of wtMpgS and its H309A mutant, the ligand lies in the catalytic pocket, located near the bottom of the central β-sheet (with reference to Fig. 2.3), and in a cleft formed between the Rossmann-like and the C-terminal regions, bordered by several loops. Here, three binding regions can be described according to Fig. 2.6 and Fig. 2.8: the typically conserved NDP binding pocket, formed by the Rossmann-like sub-domain; the sugar binding pocket, located at the boundary of the cleft region; and the less conserved acceptor binding region, located between the α-helical bundle and the C-terminal region of the central β-sheet. At the pocket bottom, and only observed in the crystal structures of apo-MpgS and wtMpgS:GDP-Man:Mg^{2+}, a second metal ion position partially occupied by Zn^{2+} lies about 6 Å below the β-face of the mannose ring (with reference to Fig. 2.8B). A number of hydrophilic residues participate in GDP-Man:Mg^{2+} recognition and binding (Fig. 2.6; Table 2.4). These are conserved in *P. horikoshii* MpgS (PDB 2ZU8), establishing a common network and scaffold environments representative of the GT55 family. The evolutionary relationship among the four diverging enzymes; MpgS, MgS, GpgS and MpgS/GpgS was assessed, outlining key residues and structural constraints relevant to catalysis (Fig. 2.7). The conservation of a global scaffold in the catalytic pocket is shown by their solvent exposed NDP-sugar binding pockets allowing ready access of the NDP-sugar. From here, as represented in Fig. 2.8A, two tunnels cross the protein bulk leading back to the surface: the first (tunnel I) lies more or less along the orientation of the NDP-sugar binding pocket, while the second (tunnel II) is roughly perpendicular to it, and begins near the bottom of the sugar ring binding region

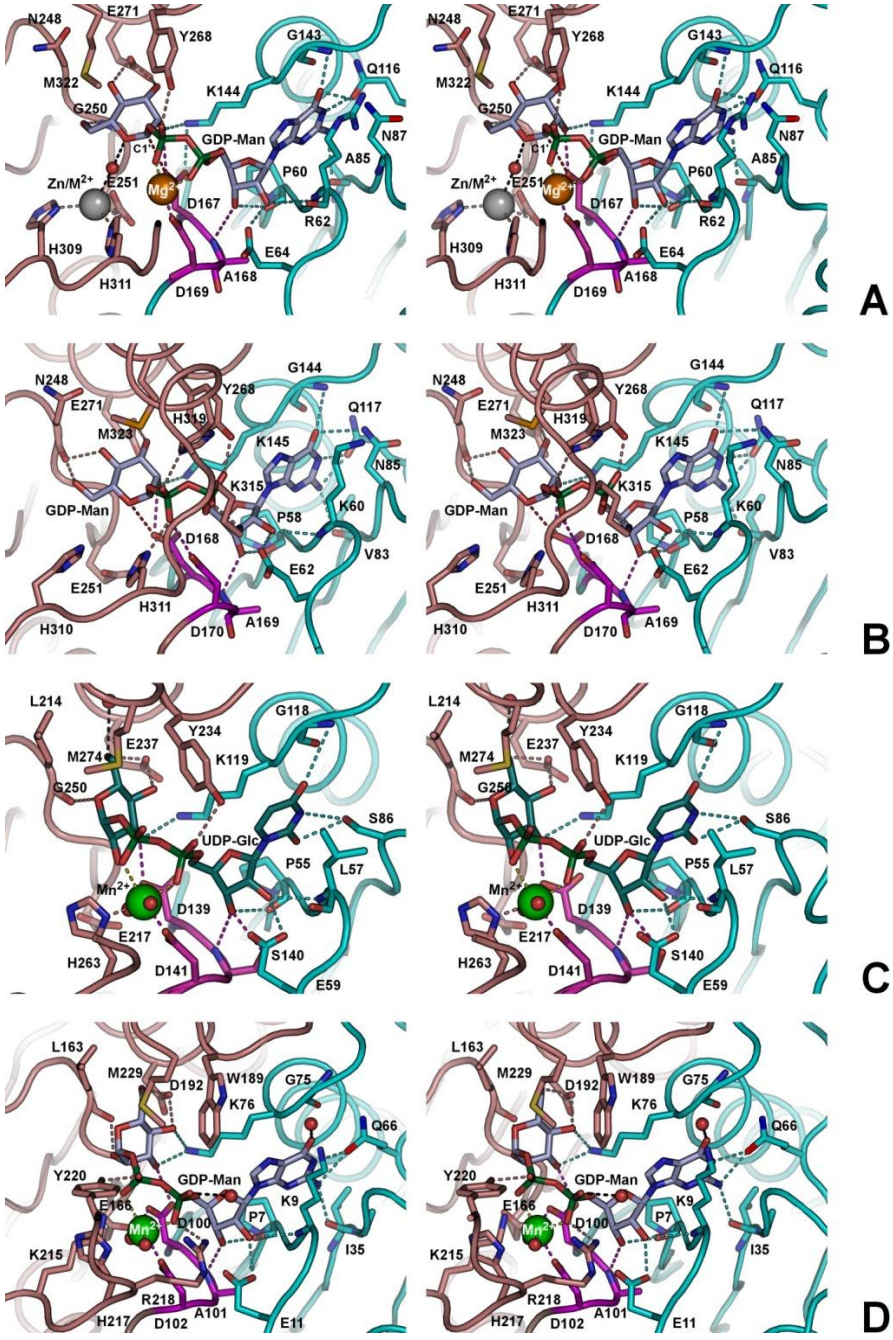

N248
E271
Y268
G143
Q116
M322
K144
G250
GDP-Man
N87
A85
Zn/M2+
E251
Mg2+
P60
D167
R62
H309
E64
H311
A168
D169
A
N248
E271
Y268
G144
Q117
H319
M323
K145
N85
K315
GDP-Man
K60
P58
D168
V83
E62
E251
H311
H310
D170
A169
B
L214
G118
Y234
E237
M274
K119
G250
S86
UDP-Glc
P55
L57
Mn2+
D139
E217
S140
H263
D141
E59
C
L163
M229
D192
W189
G75
K76
Q66
Y220
GDP-Man
E166
K9
Mn2+
D100
P7
K215
I35
R218
H217
D102
A101
E11
D

Figure 2.6 The NDP-sugar binding pocket in MpgS and structurally related enzymes. In these stereo pictures, the protein backbone is represented as a C^{α} tube diagram, the side-chains and NDP-sugar ligands are drawn as sticks, and the metal ions and the water molecule are drawn as spheres. The protein backbone and the side-chain carbon atoms are colored according to the protein regions described in the text: Rossmann-like region (cyan), DAD motif (purple) and C-terminal α-helical region (salmon). Other atom colors are blue for nitrogen, red for oxygen, yellow for sulfur, gold for selenium, green for phosphorus, orange for Mg^{2+}, green for Mn^{2+} and grey for Zn/M^{2+}. The dashed lines indicate the hydrogen bonds or electrostatic interactions established between protein side-chain atoms and the bound NDP-sugar and metal ion ligands, and are colored according to the respective protein regions. *A*, *T. thermophilus* HB 27 MpgS.; *B*, *P. horikoshii* MpgS (PDB 2ZU8). M323 is a selenomethionine. A dashed red line represents the van der Waals proximity of Asp167 to the anomeric C1` carbon in *T. thermophilus* (panel *A*), and of Asp 168 to O5´ oxygen in *P. horikoshii* (panel *B*), of the mannosyl moieties; *C*, the putative GpgS from *M. avium paratuberculosis* encoded by MAP2569c (PDB 3CKQ); *D*, *R. marinus* MgS (PDB 2BO8). Figure created with PyMol (http://www.pymol.org).

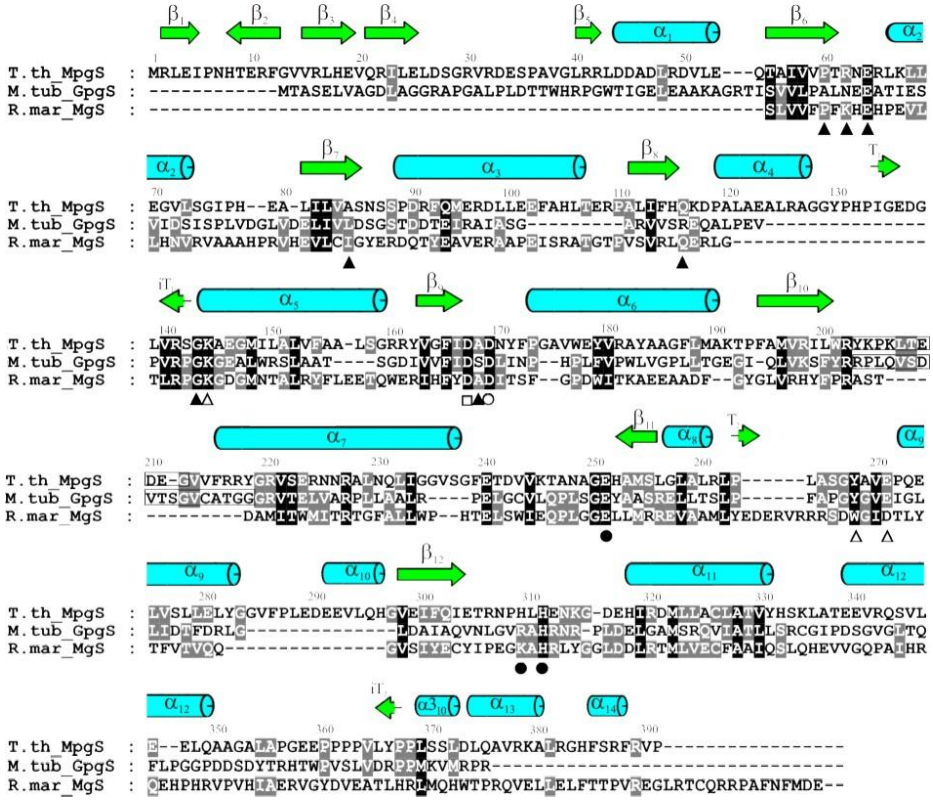


Figure 2.7 MpgS and its structural homologues. Structural alignment of *T. thermophilus* HB27 MpgS with its closest homologues representing the retaining GT81 and GT78 families: *M. tuberculosis* H37Rv GpgS (PDB 3E25) and *R. marinus* MgS (PDB 2BO8). For clarity, the other known structures from members of the GT81 and GT55 families are not included. Sequence alignment was done with ClustalW v.2.0.10 (http://www.ebi.ac.uk/Tools/clustalw2) using the default parameters and edited with GENEDOC v.2.0 in agreement with a Secondary Structure Matching calculation (*44*) in COOT. Secondary structure elements correspond to the MpgS reference structure and are labeled accordingly to the topology diagram in Fig.2.4. Helices are shown as cylinders, and open rectangles indicate the regions not modelled due to poor electron density. Conserved catalytic residues are marked accordingly to their binding specificity: for the NDP-leaving group, closed triangles; for the sugar-donor, open triangles; for the

catalytic Lewis acid divalent metal ion, open circles; for the second metal ion site in MpgS, closed circles. The open square at Asp167 denotes its proposed dual functionality in binding the mannopyranosyl moiety and in acting as a catalytic nucleophile, either in the formation of a glycosyl-enzyme intermediate according to the double-displacement S_N2 reaction mechanism, or in the stabilization of the oxocarbenium ion-like intermediate according to the D_N*A_{Nss} (S_Ni-like) reaction mechanism.

The pocket bottom is located at the boundary of the cleft region, formed by residues Pro60-Glu64 and by the DXD motif. It forms a type IV β-turn and its influence spans two moieties in the NDP-sugar substrate, since it is involved in sugar recognition and binding through Asp167, as well as in the stabilization of the ribose moiety. In addition, through the *syn*-coordination of the catalytic Lewis acid Mg^{2+} by Asp169, it is also indirectly interacting with the pyrophosphate (Figure 2.6).The significance of this motif in catalysis has been extensively studied by mutagenesis, which always resulted in an activity drop below detectable levels (*4, 6, 46*). While the second aspartate is prevalently associated with Lewis acid coordination, the first is found to interact with the sugar-moieties or their respective analogues

The β-face of the mannosyl moiety docks against the main chain of the conserved Gly250 at the back of the sugar binding pocket. The α-face is sheltered by a hydrophobic platform, which is also a topologically conserved feature among the retaining GT-A enzymes (*4*). Hydrophobic contacts are provided by Tyr268 and to a lesser extent by Met322. In both the wild-type and H309A mutant binary complexes, the mannopyranosyl-moiety was found in the typically stable $^4C_1$ chair conformation (Fig. 2.12 A and B), and with an orientation similar to that found in *P. horikoshii* MpgS (PDB 2ZU8), regardless of a Lewis acid absence in this structure. In contrast, in both MgS (PDB 2BO8) and GpgS (PDB 3CKQ), their sugar-moieties were oriented in a different binding mode (Figure 2.6).

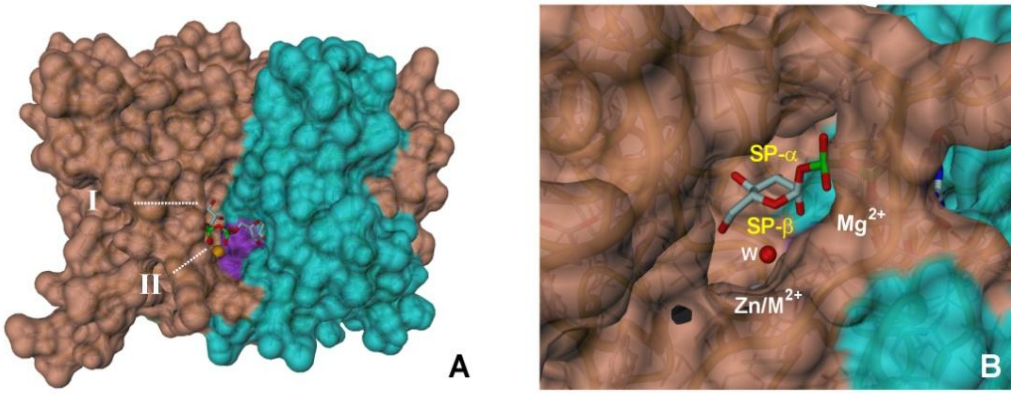


Figure 2.8 Two tunnels lead from the NDP-sugar binding pocket to the protein surface of MpgS. The solvent-accessible molecular surface was calculated with MSMS (*45*) using the coordinates of protein chain A in wtMpgS:GDP-Man:Mg^{2+} with a probe radius of 1.4 Å. The protein regions described in the text which are visible in this orientation are colored cyan (Rossmann-like region, purple (DAD motif) and salmon (C-terminal α-helical region). The protein backbone is represented as a gold-colored C^{α} tube diagram, the amino-acid residues side-chains and the ligand GDP-Man are drawn as sticks, and the metal ions and water molecules are drawn as spheres. Atom colors are light blue for carbon, blue for nitrogen, red for oxygen, green for phosphorus, orange for Mg^{2+} and grey for Zn/M^{2+}. In this figure, part of Loop1 (residues 205-215) not visible in the X-ray structure of MpgS:GDP-Man:Mg^{2+} was built using apoMpgS chain A as a template. ***A*****,** overall view showing the NDP-sugar binding pocket and the entrance to tunnels I and II; the orientation of the tunnels is indicated by white dotted lines; ***B***, detail view showing the tunnel II entrance; **SP-α**, sugar pocket α-face; **SP-β**, sugar pocket β-face. Figure created with DINO (http://www.dino3d.org).

The most notable consequence is that in MpgS Asp167 $O^{\delta 2}$ accepts a hydrogen bond from O2' and is also within van der Waals distance (*ca.* 3.5 Å) from the anomeric carbon C1', or alternatively from the O5` oxygen (*ca.* 3.1 Å) as observed for *P. horikoshii* (Fig. 2.6 A and B; Table 2.4). This environment differs from that observed in MgS and GpgS, involving the O3` and O2` oxygen atoms instead (Fig. 2.6 C and D; Table 2.4). These differences may result from different torsion angles adopted by the pyrophosphate backbone, depending on the Lewis acid metal (none, Mg^{2+} or Mn^{2+}) present. The GDP-Man orientation found in the MpgS structures has mechanistic implications, which will be discussed below.

MpgS has shown catalytic specificity for 3-PG (*10*), whereas acceptor plasticity has been reported for MgS (*4*) and to a lesser extent for GpgS (*31*). As the three enzymes share a global catalytic scaffold, 3-PG is also likely to bind to MpgS in a structurally equivalent position to that proposed for 3-PG in GpgS and D-glycerate in MgS (*4, 40, 41*). Although confirming tunnel II as the most likely entry point for 3-PG into the catalytic site, our docking calculations failed to yield a meaningful solution in terms of catalysis (Fig. 2.9 and Fig 2.10). This may be due to a limitation of the current docking methods which cannot fully account for induced fit, *i.e.*, the binding site is not rigid, but is formed during the ligand binding process. The binding mechanism of 3-PG probably acts concertedly and/or synergistically upon GDP-Man:Mg^{2+} binding, similarly to what has been shown for the inverting β-1,4-galactosyltransferase-1 and retaining α-1,3-galactosyltransferase enzymes (*7, 47, 48*). Also, optimal catalysis occurs at *ca.* 75ºC in MpgS, therefore thermally-induced protein dynamics must be considered.

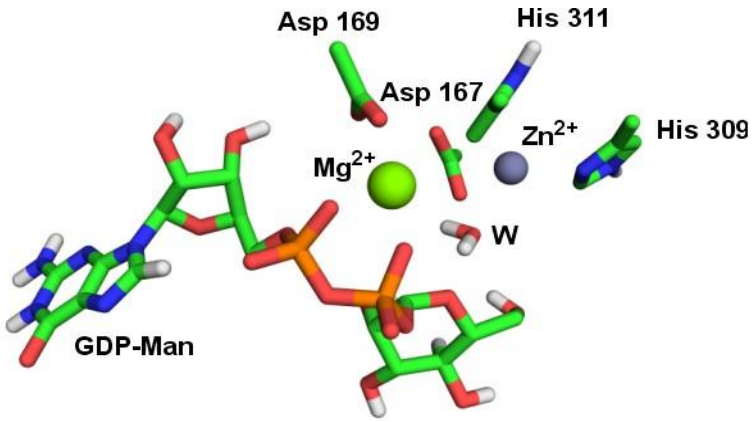


Figure 2.9 Preparing the docking calculations. Detail of the MpgS catalytic centre showing the atoms for which partial charges were calculated. Figure prepared with PyMol.

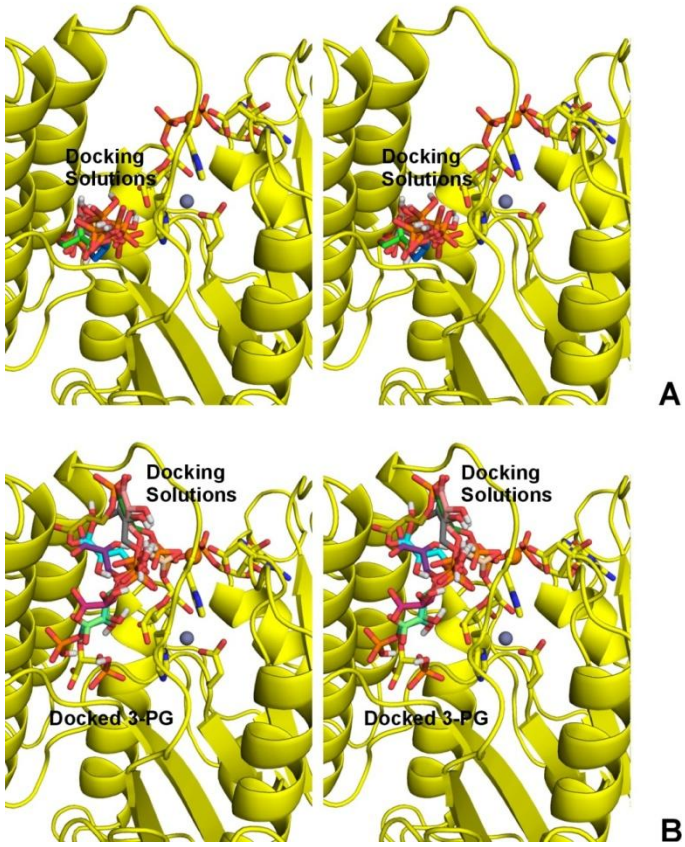


Figure 2.10 Docking results. *A*, the 10 best docking solutions of the 3-PG ligand in the catalytic pocket of wtMpgS:GDP-Man:Mg^{2+}. *B*, the 10 best docking solutions of the 3-PG ligand in the catalytic pocket of wtMpgS:GDP-Man:Mg^{2+} in the presence of a previously docked 3-PG molecule; the protein backbone is represented as a cartoon; the atoms considered to be in the MpgS catalytic centre and for which charges were calculated using quantum methods are drawn as sticks; Mg^{2+} and Zn^{2+} ions are represented as spheres, colored orange and gray, respectively; the docking solutions are represented as sticks with different colors. This figure was generated using PyMol (http://www.pymol.org).

The second metal ion coordination site

The finding of an internal Zn^{2+} site located 6Å below the β-face of the mannosyl-moiety (Fig. 2.12), led us to consider the existence of a second catalytic metal ion and its role in the glycosyl-transfer mechanism. As seen from the SHELXD results and the heavy atom parameter refinement in SHARP, this site is only partly occupied by Zn^{2+} in both monomers of the wtMpgS:GDP-Man:Mg^{2+} and apoMpgS structures. *In vivo*, a divalent metal ion other than Zn^{2+} (probably Mg^{2+} or Mn^{2+}) would be bound, and this binding site was therefore designated as Zn/M^{2+}. The metal is coordinated by Glu251 $O^{\varepsilon 1}$, His309 $N^{\delta 1}$ and His311 $N^{\varepsilon 2}$, and an axial water molecule completes the coordination sphere, although not visible in all monomers (Table 2.4). These residues belong to the sequence motif $NXGEX_nHXH$, conserved in all MpgS enzymes with known sequence.

The functional relevance of Zn^{2+} was assessed from measurements of maximal specific activity (V_{max}) (Fig 2.11). When Mg^{2+} was used as single metal co-factor at 20 mM (concentration for maximal activity) the MpgS specific activity was 45±3 μmol.min^{-1}.mg^{-1}. When measured solely in the presence of Zn^{2+} (300 μM), the activity decreased about 100-fold (0.4 ± 1.2 μmol.min^{-1}.mg^{-1}), but when 20 mM Mg^{2+} was added in combination with Zn^{2+} (300 μM) the MpgS specific activity recovered substantially (5.3 ± 0.1 μmol.min^{-1}.mg^{-1}). These findings suggest that Zn^{2+} can displace Mg^{2+} at the second metal binding site; therefore, the partial occupation of Zn^{2+} in the wtMpgS:GDP-Man:Mg^{2+} and apoMpgS structures may also be interpreted in terms of a fully occupied metal site with a mixed Zn^{2+}/Mg^{2+} population. On the other hand, the anomalous diffraction data provided no evidence for Zn^{2+} occupation at the catalytic Lewis acid position of Mg^{2+}, coordinated to GDP-Man. Kinetic assays of the MpgS mutants E251A and H309A showed negligible specific activities: none was detected for E251A and only 0.3 ± 0.2 μmol.min^{-1}.mg^{-1} were measured for H309A. These results, combined with the structural evidence for

the absence of this metal site in the H309A MpgS mutant, unequivocally show the existence of a second catalytic metal ion in MpgS.

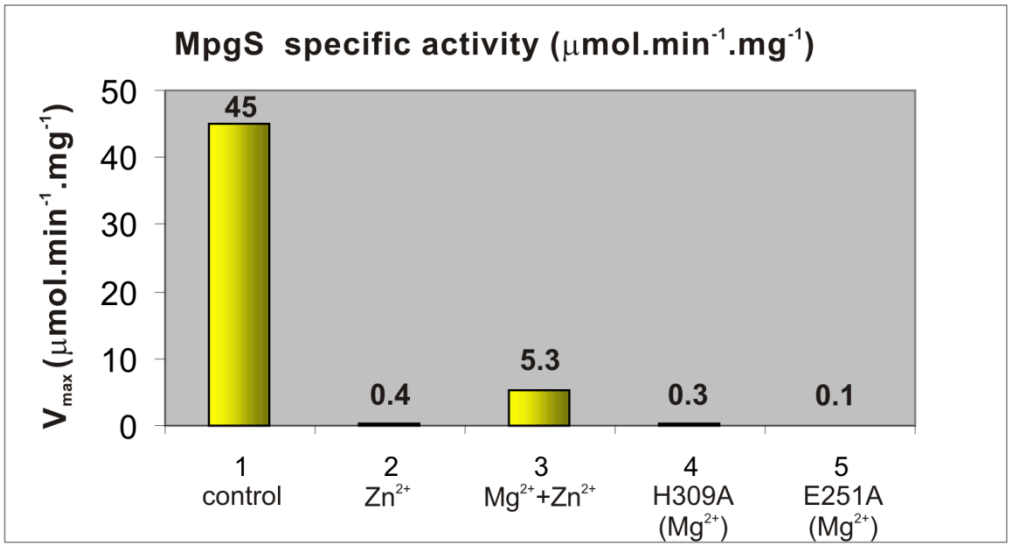


Figure 2.11 Column chart of the MpgS maximum specific activities.

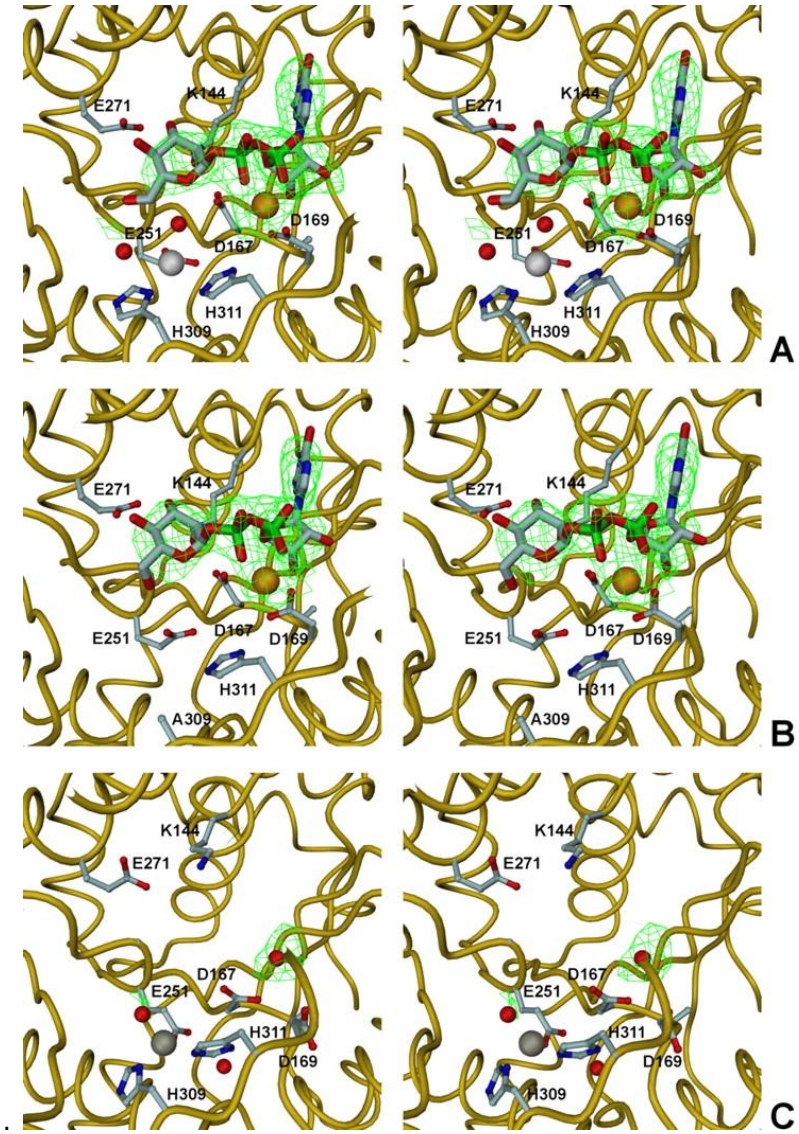


Figure 2.12 The Zn/M^{2+} binding site in the MpgS structures. Stereoviews showing the NDP-sugar binding pocket and the Zn/M^{2+} binding site in *A*, wtMpgS:GDP-Man:Mg^{2+}; *B*, H309A MpgS:GDP-Man:Mg^{2+}; *C*, apoMpgS. The protein backbone is represented as a gold-colored C$^{\alpha}$ tube diagram, GDP-Man is drawn as sticks, the side-chains of the residues mentioned in the text are labeled and drawn as ball-and-stick, the metal ions and water molecules are

drawn as spheres. Atom colors are light blue for carbon, blue for nitrogen, red for oxygen, green for phosphorus, orange for Mg^{2+} and grey for Zn/M^{2+}. In *A* and *B*, the SigmaA-weighed $|F_o|$ - $|F_c|$ electron density maps covering GDP-Man:Mg^{2+} after the first round of refinement, including only the protein chain and the Zn^{2+} ions, are represented as mesh green surfaces and drawn at the map 3 r.m.s.d. level. In *C*, a similar SigmaA- weighed $|F_o|$ - $|F_c|$ electron density map reveals only a water molecule in the same region. Figure prepared with DINO.

2.5 DISCUSSION

The conserved catalytic scaffold and its chemical significance for glycosyl transfer

The structure of *T. thermophilus* MpgS, and comparison with its closest homologues, highlights the molecular features underlying the functional specialization, namely for the synthesis of glucosylglycerate and mannosylglycerate. Despite sharing a common catalytic scaffold with other GT-A enzymes, discrete structural changes at the NDP-binding pocket are relevant for nucleoside specificity, as exemplified by the steric discrimination for UDP-Glu over GDP-Glu in GpgS (*40, 41*) provided by the local secondary structure and the orientation of the Ser81 side chain. On the other hand, acceptor specificity is more complex as it is dependent on the scaffold plasticity that accompanies catalysis, as suggested by the different conformations of the flexible regions equivalent to Loops 1 and 2 in MpgS (Figure 2.13) and by point mutations (*4*). The previously assigned structural motif at the β-face of the sugar-moiety for the retaining GT-A enzymes (*4*) is represented in MpgS by Lys144 and Glu271. In MpgS, Glu271 is located near GDP-Man, and interacts with the mannosyl-moiety (Fig. 2.6A). In fact, two charged residues at the β-face of the sugar-pocket region are topologically conserved in both inverting and retaining GTs, binding either the sugar-donor or its acceptor (Fig. 2.14).

Asn248, structurally equivalent to Gln189 in LgtC, is thought to play a central role in the D_N*A_{Nss} mechanism, where it is presumed to provide the mild nucleophilic character for the stabilization of the developing oxocarbenium-ion transition-state, without an effective covalent bond formation. However, in *T. thermophilus* and *P. horikoshii* MpgS its side chain is not properly positioned for such a role, at least in the observed conformation of the sugar-donor. Still, the role of Asn248 in catalysis should not be disregarded, as it is part of a conserved sequence motif NXGEX_nHXH present in all annotated MpgS of the GT55 family, and may play a role in acceptor binding.

The unique orientation of the sugar-moiety found in both MpgS structures places its anomeric carbon in close proximity to the conserved Asp167 of the DXD motif (Fig. 2.14A). Such orientation was independent of the presence of the catalytic metal, and may be representative of the GT55 family since it was also seen in the structure of *P. horikoshii* MpgS, despite the absence of a catalytic Lewis acid. Thus, it may be speculated that different strategies for oxocarbenium-ion stabilization may exist, leading to alternative pathways with the preservation of the net reaction stereochemistry. Herein we propose a more complete structural signature that tailors the stereo-specific outcome, comprising Asp167-Lys144-Glu271, combined with the orientation (*in* vs. *out*) of the glycosidic bond linking the β-phosphate to the anomeric carbon relative to the DXD motif, as summarized in Table 2.5, the mechanistic implications of which are discussed below.

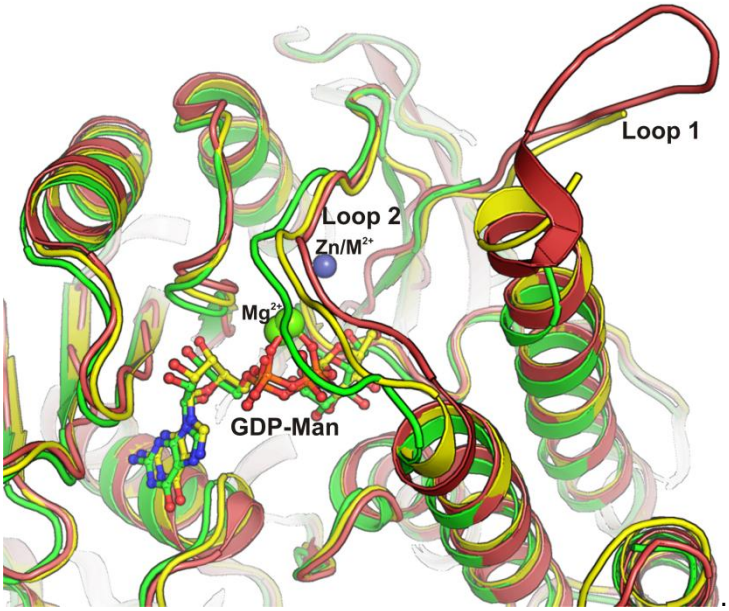


Fig. 2.13 Detail of the alternative conformations adopted by Loop 1 and Loop 2. Cartoon representation of the superposed structures wtMpgS: GDP-Man:Mg^{2+} structure (yellow; chain B), apo-MpgS (salmon; chain A) and H309A MpgS:GDP-Man:Mg^{2+} (green; chain B). The exit site of tunnel 2 by which 3-PG may access the catalytic active site, is shaped by the interface of loops 1 and 2. Ligand GDP-Man is represented in "ball-and-stick" and metal ions are show in spheres (green for Mg^{2+} and dark gray for Zn^{2+}. Figure prepared with PyMol (http://www.pymol.org).

Asp167 as the putative catalytic nucleophile in the retaining catalysis

In contrast with the relative high specificity observed for the NDP group (Fig. 2.6), the sugar pocket dimensions allow for binding of other monosaccharyl-moieties, as well as changes in orientation for any given sugar ring. The natural availability of these moieties is a restricting factor (*49*) and catalytic proficiency is likely to require a well-defined sugar-ring stereochemistry, as well as a proper orientation of its scissile glycosidic bond towards the incoming proton-donor, in a similar fashion to the stereospecific

protonation in GHs (*50*). In the retaining GTs, the incoming acceptor is likely to act as a Brønsted acid, prior to its nucleophilic attack at the anomeric carbon center (Fig 1.4) (*5*). In the MpgS structures, a lone pair of O^{1B} oxygen in GDP-Man is efficiently oriented towards the entrance of Tunnel II, the expected entry point of 3-PG (Figure 2.14). Its C2(OH) group, once activated, attacks the mannose anomeric carbon C1' by nucleophilic substitution, either via the double-displacement S_N2-like mechanism, or by the $D_N{*}A_{Nss}$ mechanism (Fig. 2.14). In general, the breaking of the glycosidic bond occurs together with the formation of several oxocarbenium ion-like transition states and short-lived intermediaries (*5*). In the MpgS structures, this charge development can be stabilized by Asp167 at C1` in *T. thermophilus* or alternatively at O5` in *P. horikoshii*. This residue is hydrogen-bonded to O2' in the mannopyranosyl ring as well as to N^{ζ} of the conserved Lys144 and these interactions effectively fix the side-chain conformation of Asp167 in a suitable orientation to act either as a nucleophilic stabilizer in the retaining $D_N{*}A_{Nss}$ mechanism, or as an effective catalytic nucleophile according to the double-displacement S_N2-like mechanism. In the latter case, the formation of a glycosyl-enzyme intermediate would imply a reorientation of the C2`(OH) epimer (most likely to a pseudoequatorial position) as the sugar-puckering evolves into a transition state conformation analogous to that proposed for several mannosidases (*51*).
Despite sharing a common scaffold with MpgS, the equivalent Asp residue in MgS and GpgS is hydrogen-bonded to O3' instead (Fig. 2.6 C and D), and it can be argued that in those crystal structures, the sugar donor is not in an orientation leading to efficient catalysis. This may have been an artifact of co-crystallization with Mn^{2+}.[1] However, the structural changes that very likely accompany entry of the acceptor into the active site pocket may drive the sugar ring back into an orientation suitable to catalysis according to one of the two proposed mechanisms.

An analysis of the orientation of the scissile bond in relation to the DXD motif found in both inverting and retaining GT-A enzymes showed that, regardless the nature of the co-catalytic metal ion, the *in* orientation is

prevalent in the former while in the latter it was mostly observed in the *out* orientation (Table 2.5). The *in* orientation favors the presence of an internal proton donor, whereas the *out* orientation satisfies the stereochemistry required for an in-line protonation from the sugar acceptor. For example, in the inverting β-1,4-galactosyltransferase-1 (Gal-T1) the β-phosphate was shown to be in the *in* orientation, properly placed to be protonated by a conserved water molecule that interacts with the DXD motif and the metal-Lewis acid (Fig. 2.14B). Such protonation is thought to be mandatory for the irreversibility of the reaction, by preventing the re-forming of the glycosidic bond, and thus facilitating the nucleophilic attack from the incoming acceptor at the β-face (*48*). Alternatively, our results suggest that, at least for some GT-A retaining enzymes, Asp167 in the DXD motif can work as the catalytic nucleophile.

A second catalytic metal ion in MpgS enzymes

In addition to the Lewis acid metal coordinated to GDP-Man, found in the *T. thermophilus* MpgS:GDP-Man:Mg^{2+} binary complex, our mutagenesis studies highlighted the requirement of a second catalytic metal ion for efficient catalysis in *T. thermophilus* MpgS. The coordination site is formed by Glu251, His309, His311 and an axial water molecule. Equivalent residues are present in *P. horikoshii* MpgS, however no metal was found in any of its structures (PDB 2ZU7, 2ZU8 and 2ZU9). This can be explained by the absence of any metal-ion with chemistry equivalent to Zn^{2+} in the crystallization buffer, contrary to *T. thermophilus* MpgS

[1] Recently new structures of *R. marinus*` MgS in complex with GDP-Man: Mg^{2+} and with GDP-Man:Mn^{2+} reported by Nielsen et al (*52*), revealed the same binding mode for the sugar moiety as that previously reported by Flint and co-workers (*4*) and herein described, suggesting that other factors may be involved for such preferred binding mode in these structures (see text).

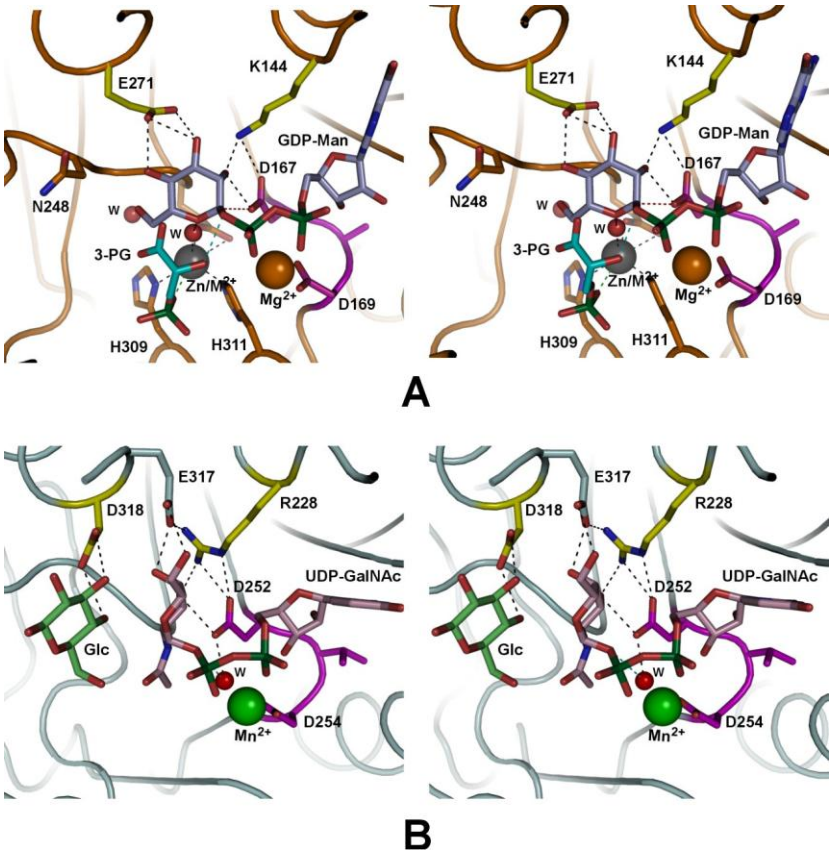


Figure 2.14 Structural signatures in retaining and inverting glycosyltransferases with GT-A fold. A comparison of the structural signatures present in two representative structures of the retaining and inverting GTs with GT-A fold, bound to their sugar-donor**.** *A*, the predicted catalytic complex of *T. thermophilus* MpgS. The scissile glycosidic bond is oriented in the *out* configuration w.r.t the DXD motif favoring an *in-line* protonation (cyan dashed line) from the incoming 3-PG, prior to its nucleophilic attack at the anomeric carbon center C1'. *B***,** merged coordinates of the inverting Gal-T1 with its UDP-GalNAc donor (from PDB 1OQM) and Glc acceptor (from PDB 2FYD) oriented as MpgS. The scissile bond adopts the *in* configuration positioning the glycosidic oxygen to be protonated by the structural water molecule (W; cyan dashed line). The protein backbones are

represented as C^{α} tube diagrams, bright orange for MpgS and light cyan for Gal-T1. The DXD motif is colored magenta in both molecules. The substrate molecules and the side chains of labeled protein residues are drawn as sticks, and the metal ions and water molecules are represented by spheres. Carbon atoms are colored yellow in the signature residues, magenta in the DXD motif, light blue in GDP-Man, cyan in 3-PG, light pink in UDP-GalNAc and light green in Glc. Other atoms are colored red for oxygen, blue for nitrogen, dark green for phosphorus, orange for Mg^{2+}, grey for Zn/M^{2+} and green for Mn^{2+}. Hydrogen bonds are represented as dashed lines.

The coordinating His309 is replaced by a Lysine in MgS, and by an Arginine in both GpgS and the MpgS/GpgS from *R. xylanophilus.* These residues are less likely to act as metal-ion ligands, but their positively charged side chains may play a role parallel to that of the second catalytic metal ion in *T. thermophilus* MpgS. Elimination of the Zn/M^{2+} binding site via an E251A or H309A mutation effectively destroyed enzyme activity, by abolishing the metal-coordination ability, as seen in the H309A structure. This effect can be rationalized in terms of an interaction between one of the negatively charged phosphate oxygens of 3-PG with the positively charged second catalytic metal ion. Indeed, a simple manual modeling of a 3-PG molecule in its expected binding region shows that not only this interaction is possible, but also the C2(OH) group becomes oriented towards the glycosidic oxygen (less than 3Å distant), as expected during general acid catalysis (Fig. 2.14A). Therefore, without this metal centre, the inability for 3-PG binding in the correct orientation will hinder catalysis. This notion is supported by the structure of the putative GpgS from *M. avium paratuberculosis* (PDB 3CKJ), where a citrate molecule occupies a similar position to 3-PG as in Fig.2.14, and interacts with Arg261 through its γ-carboxylate group. This binding position could not be reproduced by our docking calculation since it requires a local rearrangement of the protein structure to eliminate close contacts (2.5 - 3 Å) with 3-PG.

Zn^{2+} was shown to act as an inhibitor of MPG synthesis, competing with Mg^{2+} in the occupation of the second metal ion binding site. Since the crystallization buffer contained 0.4 - 0.6 mM Zn^{2+} and 200 mM Mg^{2+}, the affinity of the second metal binding site for Zn^{2+} is clearly much higher than that for Mg^{2+}. The inhibitory effect of Zn^{2+} may have two origins: (1) due to its acidity, which, via its axial water molecule destabilizes the oxocarbenium-ion transition state, similarly to the 5-fluoro-sugar inhibitors, or (2) to a higher affinity for one of the reaction products, preventing its release from the catalytic pocket and therefore slowing down the reaction turnover. Also, under physiological conditions the second metal ion binding site may be labile. Indeed, the degree of structural flexibility in this region comprised by flexible loops 1 and 2 may transiently shape this site for binding a second metal ion during the catalytic cycle (Fig. 2.13). In particular, His311 in *T. thermophilus* MpgS is positioned in the vicinity of Loop2 and may coordinate either one of the two catalytic metal ions. This is reinforced by the alternate conformations of the equivalent residue His312 in two *P. horikoshii* structures, interacting with either GDP-Man (PDB 2ZU8; Figure 2.6B) or GDP:Co^{2+} (PDB 2ZU9). In the latter structure, the imidazole group of His312 moves away from the second catalytic metal site and coordinates instead the Lewis acid metal of the GDP:Co^{2+} leaving group, possibly destabilizing the second catalytic metal centre, leading to the release of its bound metal ion. This release most likely follows that of the MPG product, while the GDP:M^{2+} leaving group is stabilized by residues equivalent to His311 in *T. thermophilus* MpgS.

Table 2.1 - Data collection statistics for the three MpgS crystals

	wtMpgS:GDP-Man:Mg^{2+} MAD data set			wtMpgS: GDP-Man:Mg^{2+}	H309A MpgS: GDP-Man:Mg^{2+}	apo-MpgS
	Peak	Inflection	Remote			
Beamline		ESRF ID 29		ESRF ID 23-1	DLS I04	ID14-4
Detector			ADSC Quantum Q315r			
Wavelength (Å)	1.2823	1.2831	1.2716	1.0672	0.9763	1.3050
Data Processing			XDS			
Space Group			$P\ 4_1\ 2_1\ 2$			
Unit cell parameters (Å)	a=113.1, c=197.1	a=113.5, c=197.7	a=114.2, c=198.0	a=113.7, c=196.7	a=112.7, c=199.7	a = 113.3, c = 197.1
Resolution (Å)	45.18-2.99 (3.17-2.99)	45.31-2.99 (3.17-2.99)	45.41-2.99 (3.17-2.99)	45.1-2.80 (2.97-2.80)	48.90-2.98 (3.16-2.98)	45.2-2.97 (3.10-2.97)
Nr. Observations	481922 (75644)	486946 (74381)	196686 (21481)	197746 (31785)	69101 (7084)	507641 (75956)
Unique reflections[a]	49479 (7960)	49862 (7861)	50076 (7601)	32332 (5078)	25598 (3441)	26864 (4114)
Completeness (%)	99.8 (99.4)	99.6 (97.8)	98.6 (92.6)	99.3 (98.8)	94.0 (80.6)	99.4 (96.7)
R-merge (%) [b]	9.4 (216.8)	10.1 (217.1)	10.4 (164.9)	6.8 (151.1)	4.5 (37.2)	10.8 (127.2)
$\langle I/\sigma(I)\rangle$	16.18 (1.29)	14.90 (1.26)	8.33 (0.61)	15.62 (1.23)	15.41 (2.09)	19.5 (2.18)
R-meas (%) [c]	10 (229.2)	10.7 (229.4)	12.0 (198.9)	7.5 (165.1)	5.6 (47.9)	11.1 (130.8)
Z [d,e]	2					
Estimated Solvent Content (%) [e]	65.7					

[a] In the MAD dataset, Bijvoet pairs were treated as separate observations during the scaling process. [b] R-merge = merging R-factor, $(\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|) / (\sum_{hkl} \sum_i I(hkl)) \times 100\%$.; [c] R-meas = redundancy independent R-factor (intensities), $\sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl) \times 100\%$. (*53*); [d] Predicted nr. molecules in the asymmetric unit; [e] from Matthews coefficient calculations (*54*)

Table 2.2 - Phase refinement statistics for wtMpgS:GDP-Man:Mg^{2+}

	"Native"	Peak	Inflection	Remote
ESRF Beamline	ID 23-1		ID 29	
Wavelength (Å)	1.0672	1.2823	1.2831	1.2716
FOM, acentric			0.453 /0.437	
FOM, centric			0.454 /0.449	
Phasing power, acentric [a]	0.755 /0.705	1.173 /1.131	1.121 /1.080	-
Phasing power, centric [a]	0.459 /0.445	0.754 /0.749	0.642 /0.628	-
Phasing power, anomalous [a]	0.348 /0.296	0.868 /0.703	0.621 /0.522	0.441/0.389
Acentric R_{cullis} [a]	0.618 /0.604	0.681 /0.709	0.663 /0.659	-
Centric R_{cullis} [a]	0.692 /0.677	0.947 /0.974	1.050 /1.057	-
Anomalous R_{cullis} [a]	0.970 /0.977	0.871 /0.903	0.928 /0.944	0.965/0.971
Overall $\|E^2\|$ correlation [a,b]			0.521 /0.432	
FOM after final DM run [a,b]			0.699 /0.851	
FOM after NCS averaging [a,c]			0.799 /0.884	

[a] First figure refers to SHARP results with Zn^{2+} isotropic B-factor refinement, second figure to SHARP results with Zn^{2+} anisotropic B-factor refinement; [b] from SHARP density modification procedure (SOLOMON followed by final DM run) (*17*);[c] from DM (*18*).

Table 2.3 - Final refinement statistics for the three MpgS crystal structures

	wild-type apoMpgS		wild-type MpgS:GDP-Man:Mg^{2+}		H309A MpgS:GDP-Man:Mg^{2+}	
Resolution limits (Å)	42.9 - 2.97 (3.05 - 2.97)		45.1 - 2.81 (2.88 - 2.81)		98.2 - 2.98 (3.05 - 2.98)	
R-factor (%) [a]	0.190 (0.315)		0.176 (0.339)		0.174 (0.261)	
nr. working reflections, no $\sigma(F_o)$ cutoff	25502 (1558)		30259 (2142)		24263 (1155)	
Free R-factor (%) [b]	0.256 (0.369)		0.234 (0.426)		0.232 (0.299)	
nr. free reflections, no $\sigma(F_o)$ cutoff	1358 (81)		1620 (115)		1292 (51)	
Overall coordinate error estimate (Å) [c]	0.32		0.27		0.27	
Model composition						
non-hydrogen protein atoms	6076		6025		5881	
GDP-Man	-		78		78	
Zn^{2+}	5		5		3	
Mg^{2+}	-		2		2	
citrate	13		-		-	
Solvent molecules	80		36		15	
Model r.m.s. deviations from ideality						
Bond lengths (Å)	0.015		0.016		0.016	
Bond angles (º)	1.7		1.7		1.6	
Chiral centers ($Å^3$)	0.10		0.10		0.11	
Planar groups (Å)	0.008		0.007		0.008	
Model completeness and validation	Chain A	Chain B	Chain A	Chain B	Chain A	Chain B
Regions omitted	-	203-211	206-211	203-211	203-212 352-357	203-215 352-357
Mean B values ($Å^2$) [d]						
protein main-chain	107	102	108	105	89.9	85.2
protein side-chain	110	105	114	112	94.8	89.7
GDP-Man	-		115	108	85.1	88.4
Mg^{2+}, Zn^{2+}	88.9		99.9		71.8	
citrate	106.2		-		-	
solvent	65.6		64.9		47.0	
Ramachandran plot statistics. Residues in:	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*
most favored regions (%)	85.7	85.7	87.9	87.5	90.7	91.5
allowed regions (%)	13.1	13.1	11.2	11.3	8.4	8.5
generously allowed regions (%)	0.9	0.9	0.9	1.2	0.6	0
disallowed regions (%)	0.3	0.3	0.0	0.0	0.2	0
PROCHECK G-factor	-0.15	-0.19	-0.10	-0.08	-0.05	-0.03

[a] R-factor = $\sum_{hkl} ||F_o|-|F_c|| / \sum_{hkl} |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively; [b] Free R-factor is the cross-validation R-factor computed from a randomly chosen subset of 5% of the total number of reflections, which were not used during the refinement. The same set of hkl indices was used for the three structures; [c] Maximum-likelihood estimate with REFMAC; [d] Calculated from equivalent isotropic B values, including the TLS contribution for the protein and GDP-Man atoms.

Table 2.4 - Hydrogen bond and metal ion coordination distances in the MpgS NDP-sugar binding pocket

NDP/ion atom	acceptor atom	distance (Å) [a]	acceptor in *P.h.* MpgS [b]	distance (Å) [a]	acceptor in GpgS [c]	acceptor in MgS [d]
nucleoside						
N1	Gln116 $O^{\varepsilon 1}$	2.5 (2.6)	Gln117 $O^{\varepsilon 1}$	2.7 (2.6)	- [e]	Gln66 $O^{\varepsilon 1}$ [f]
N2	Ala85 O	3.2 (3.2)	Ala83 O	3.1 (2.8)	-	Ile35 O
	Gln116 $O^{\varepsilon 1}$	3.0 (2.9)	Gln117 $O^{\varepsilon 1}$	2.8 (2.9)	-	Gln66 $O^{\varepsilon 1}$
O6	Gly143 N	2.7 (2.7)	Gly144 N	3.1 (-)	Gly118 N (>3.2Å)	W
	Ans87 $O^{\delta 1}$	3.1 (3.2)	Asn85 $N^{\delta 2}$	3.1 (3.1)	-	-
O2*	Arg62 N	3.0 (3.1)	Lys60 N	3.0 (3.0)	Leu57 N	Lys9 N
	Glu64 $O^{\varepsilon 1}$	2.7 (2.5)	Glu62 $O^{\varepsilon 2}$	2.6 (2.8)	Glu59 $O^{\varepsilon 1}$	Glu 11 $O^{\varepsilon 1}$
O3*	Pro60 O	3.0 (2.7)	Pro58 O	3.0 (2.9)	Pro55 O	Pro7 O
	Ala168 N	3.2 (3.3)	Ala169 N	2.9 (3.0)	Ser140 N,O^{γ}	Ala101 N
diphosphate [g]						
O1A	-	-	Tyr268 O^{η}	2.4 (2.4)	Tyr234 O^{η}	W
O2A	-	-	Asp170 $O^{\delta 2}$	3.0 (3.1)	W	-
	-	-	Lys315 N^{ζ}	2.5 (-)	-	-
O2B	Tyr268 O^{η}	2.8 (3.1)	His319 $N^{\varepsilon 2}$	2.9 (3.1)	-	Tyr220 O^{η}
O3B	-	-	Glu170 $O^{\delta 2}$	3.0 (2.9)	W	-
	-	-	His312 $N^{\delta 1}$	2.6 (2.7	-	-
sugar						
O2'	Asp167 $O^{\delta 1}$	2.9 (2.6)	Asp168 $O^{\delta 2}$	3.1 (2.6)	W	Leu163 O
	Lys144 N^{ζ}	2.6 (3.0)	Lys 145 N^{ζ}	3.1 (-)	Mn^{2+}	-
O3'	Glu271 $O^{\varepsilon 1}$	2.8 (2.9)	Glu271 $O^{\varepsilon 1}$	3.1 (2.9)	Lys114 N^{ζ}	Lys76 N^{ζ}
	Glu271 $O^{\varepsilon 2}$	3.0 (3.2)	-	-	Asp139 $O^{\delta 1}$	Asp100 $O^{\delta 1}$
O4'	Glu271 $O^{\varepsilon 1}$	2.5 (2.7)	Asn248 $O^{\delta 1}$	- (2.4)	Glu237 $O^{\varepsilon 1}$	Asp192 $O^{\delta 1}$
	-	-	Asn248 O	2.9 (2.7)	Lys119 N^{ζ}	Lys76 N^{ζ}
O5'	W1	2.7 (3.0)	Asp168 $O^{\delta 2}$	3.1 (-)	W	-
O6'	Gly250 N	2.6 (2.7)	Asn248 O	2.4 (3.1)	Glu271 $O^{\varepsilon 2}$	Asp192 $O^{\delta 1}$
metal ions						
Mg^{2+} [h]	GDP-Man O2A	1.9 (2.0)	-	-	UDP-Glc O1A	GDP-Man O2A
	GDP-Man O3B	1.9 (1.9)	-	-	UDP-Glc O2B	GDP-Man O3B
	Asp169 $O^{\delta 2}$	2.0 (1.9)	-	-	Asp141 $O^{\delta 2}$	Asp102 $O^{\delta 2}$

	-	-	-	-	His263 $N^{\delta 1}$	His217 $N^{\delta 1}$
Zn/M^{2+}	Glu251 $O^{\varepsilon 1}$	2.2 (2.4)	-	-	2 W	2 W
	Glu251 $O^{\varepsilon 2}$	3.0 (2.9)	-	-	-	-
	W1	2.7 (2.4)	-	-	-	-
	His309$N^{\delta 1}$	2.5 (2.8)	-	-	-	-
	His311$N^{\varepsilon 2}$	2.2 (2.3)	-	-	-	-

a Distances are listed first for monomer A then for monomer B (in parenthesis); distance cut-off is 3.2 Å in both monomers; **b** *P. horikoshii* MpgS:GDP-Man (PDB 2ZU8); **c** MAP2569c GpgS:UDP-Glc:Mn^{2+} (PDB 3CKQ); **d** *R. marinus* MgS:GDP-Man:Mn^{2+} (PDB 2BO8). UDP-Glc, UDP-α-D-glucose, GDP-Man, GDP-α-D-mannose; W, water molecule where W1 corresponds to the axial water in the co-catalytic metal center; **e** the uracyl N3 and O2 are hydrogen-bonded to Ser86; **f** after 180º side-chain flip in COOT; g diphosphate numbering as in *T. thermophilus* MpgS:GDP-Man:Mg^{2+}; **h** Mn^{2+} in GpgS and MgS.

Table 2.5. Representative structures of inverting and retaining classes of glycosyltransferases with GT-A fold.

Enzyme (Source)	Family	PDB	GT Class [a]	DXD	Sugar donor or analogue	Co-catalytic metal	Acceptor or analogue	Proposed nucleophile	Structure (Å)	scissile oxygen [b]	Reference (source)
α-GalT (Bo) [c]	GT6	1G93	ret	D225-A-D227	UDP-2F-Gal	Mn^{2+}	GAL	E317 [d]	2.4	*in*	(*7*)
LgtC (Nm)	GT8	1GA8	ret	D103-I-D105	UDP-2F-Gal	Mn^{2+}	DEL	Q189 [e]	2.0	*out*	(*6*)
Glycogenin 1 (Oc)	GT8	1LL2	ret	D102-A-D104	UDP-Glc	Mn^{2+}	-	D163 [d]	1.9	*out*	(*5, 46*)
MpgS (Tt)	GT55	2WVL	ret	D167-A-D169	GDP-Man	Mg^{2+}	-	D167 [e]	2.8	*out*	this work
MpgS (Ph)	GT55	2ZU8	ret	D168-A-D170	GDP-Man	Mn^{2+}	-	-	2.4	*out*	Kawamura, 2009, unpublished
MgS (Rm)	GT78	2BO8 2Y4M	ret	D100-A-D102	GDP-Man	Mn^{2+} Mg^{2+}	-	L163 [e]	2.8	*out*	(*4, 52*)
GpgS (Map)	GT81	3CKQ	ret	D139-S-D141	GDP-Man	Mn^{2+}	-	L214 [e]	3.0	*out*	(*41*)
GnT1 (Oc)	GT13	1FOA	inv	E211-D-D213	UDP-GlcNAc	Mn^{2+}	-	-	1.8	*in*	(*55*)
GlcAT-I (Hs)	GT43	1KWS	inv	D194-D-D196	UDP-GlcU	Mn^{2+}	-	-	2.1	*in*	(*56*)
Gal-T1 (Bo)	GT7	1OQM	inv	D252-V-D254	UDP-GalNAc	Mn^{2+}	-	-	2.3	*in*	(*47*)

[a] ret - retaining GT; inv - inverting GT; [b]orientation of the electron lone pairs in the glycosidic oxygen (linking the β-phosphate with the anomeric C1´ carbon) with respect to the DXD motif: *in* - one pair towards DXD, *out* - the other pair towards the acceptor; [c] behaves as an inverting GT in the presence of small nucleophile molecules, like azide and water (*7*); [d] in double-displacement S_N2-like mechanism; [e] in front-face $D_N{*}A_{Nss}$ mechanism. **Abbreviations: α-GalT:** α-1,3-Galactosyltransferase; **LgtC:** α-1,4-galactosyl transferase, **Gal-T1:** β-1,4-galactosyltransferase-1, **GpgS:** glucosyl-3-phosphoglycerate synthase, **MpgS:** Mannosyl-3-phosphoglycerate synthase, **GnT1:** N-acetylglucosaminyltransferase, 1, **GlcAT-I**: ß-1,3-glucoronyltransferase**, Gal-T1:** β-1,4-galactosyltransferase, **GDP-Man:** GDP-α-D-mannose, **UDP-2F-Gal:** UDP-2-deoxy-2-fluorogalactose, **UDP-Glc:** UDP-glucose; **UDP-GlcNAc**: UDP-N-acetylglucosamine, **UDP-GalNAc**: UDP-N-acetylgalactosamine, **UDP-GlcU:** UDP-glucoronate, **GAL:** β-D-galactose, **DEL:** 4-deoxygalactose, **Bo:** *Bos taurus*; **Nm:** *Neisseria meningitidis*; **Tt:** *Thermus thermophilus*, **Ph***:* *Pyrococcus horikoshii*; **Oc:** *Orytolagus cuniculus*, **Rm:** *Rhodothermus marinus*, **Map:** *Mycobacterium avium paratuberculosis*, **Hs:** *Homo sapiens.*

2.6 REFERENCES

1. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic Acids Res 37*, D233-238.
2. Whithers, S. G., and Aebersold, R. (1995) Approaches to labeling and identification of active site residues in glycosidases, *Protein Sci 4*, 361-372.
3. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst, *Acc Chem Res 33*, 11-18.
4. Flint, J., Taylor, E., Yang, M., Bolam, D. N., Tailford, L. E., Martinez-Fleites, C., Dodson, E. J., Davis, B. G., Gilbert, H. J., and Davies, G. J. (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase, *Nat Struct Mol Biol 12*, 608-614.
5. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annu Rev Biochem 77*, 521-555.
6. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs, *Nat Struct Biol 8*, 166-175.
7. Jamaluddin, H., Tumbale, P., Withers, S. G., Acharya, K. R., and Brew, K. (2007) Conformational changes induced by binding UDP-2F-galactose to alpha-1,3 galactosyltransferase- implications for catalysis, *J Mol Biol 369*, 1270-1281.
8. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases, *J Mol Biol 328*, 307-317.
9. Lairson, L. L., and Withers, S. G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes, *Chem Commun* (Camb, U K), 2243-2248.

10. Empadinhas, N., Albuquerque, L., Henne, A., Santos, H., and da Costa, M. S. (2003) The bacterium Thermus thermophilus, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate, *Appl Environ Microbiol 69*, 3272-3279.
11. Bradford, M. (1976) A rapid and sensitive for the quantitation of micogram quantities of protein utilizing the principle of protein-dye binding., *Anal Biochem 72*, 248-254.
12. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants, *J. Appl. Cryst. 26*, 795-800.
13. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs, *J. Appl. Cryst. 37*, 843-844.
14. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Cryst D 58*, 1772-1779.
15. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE, *Z. Kristallogr. 217*, 644-650.
16. La Fortelle, E. d., and Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for the Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods, *276*, 472-494.
17. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase, *Acta Cryst D 52*, 30-42.
18. Cowtan, K. (1994) 'dm': An automated procedure for phase improvement by density modification. , In *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, pp 34-38.
19. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Cryst D 59*, 1131-1137.

20. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta Cryst D 57*, 122-133.
21. Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors., *Acta. Crystallogr., Sect A: Found Crystallogr 32*, 922-923.
22. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.
23. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta Cryst D 62*, 1002-1011.
24. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Cryst D 60*, 2126-2132.
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst D 53*, 240-255.
26. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions, *Acta Cryst D 60*, 432-438.
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK - a program to check the stereochemical quality of proteins., *J Appl Crystallogr 26*, 283-291.
28. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF-a program to identify and analyze structural motifs in proteins, *Protein Sci 5*, 212-220.
29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res 28*, 235-242.
30. Borges, N., Marugg, J. D., Empadinhas, N., da Costa, M. S., and Santos, H. (2004) Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of *Rhodothermus marinus*, *J Biol Chem 279*, 9892-9898.
31. Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., and da Costa, M. S. (2008) Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase, *FEMS Microbiol Lett 280*, 195-202.

32. Ames, B. N. (1966) Assay of Inorganic Phosphate, Total Phosphate and Phosphatases, *Methods Enzymol. 8*, 115-118.
33. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J., J. A., Vreven, T. K., K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. (2004) *G03a: GAUSSIAN03, Revision D.01*, Gaussian, Inc., Wallingford CT.
34. Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (2002) *J. Phys. Chem. 97*, 10269-10280.
35. Stewart, J. J. P. (1989) *J. Comput. Chem. 10*, 209-220.
36. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J Comput Chem 30*, 2785-2791.
37. Richardson, J. S. (1977) beta-Sheet topology and the relatedness of proteins, *Nature 268*, 495-500.
38. Auld, D. S. (2001) Zinc coordination sphere in biochemical zinc sites, *Biometals 14*, 271-313.

39. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3, *Bioinformatics 24*, 2780-2781.
40. Pereira, P. J., Empadinhas, N., Albuquerque, L., Sa-Moura, B., da Costa, M. S., and Macedo-Ribeiro, S. (2008) Mycobacterium tuberculosis glucosyl-3-phosphoglycerate synthase: structure of a key enzyme in methylglucose lipopolysaccharide biosynthesis, *PLoS ONE 3*, e3748.
41. Fulton, Z., McAlister, A., Wilce, M. C., Brammananth, R., Zaker-Tabrizi, L., Perugini, M. A., Bottomley, S. P., Coppel, R. L., Crellin, P. K., Rossjohn, J., and Beddoe, T. (2008) Crystal structure of a UDP-glucose-specific glycosyltransferase from a Mycobacterium species, *J Biol Chem 283*, 27881-27890.
42. Sa-Moura, B., Albuquerque, L., Empadinhas, N., da Costa, M. S., Pereira, P. J., and Macedo-Ribeiro, S. (2008) Crystallization and preliminary crystallographic analysis of mannosyl-3-phosphoglycerate synthase from Rubrobacter xylanophilus, *Acta Cryst F 64*, 760-763.
43. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J Mol Biol 247*, 536-540.
44. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.
45. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) Reduced surface: an efficient way to compute molecular surfaces, *Biopolymers 38*, 305-320.
46. Gibbons, B. J., Roach, P. J., and Hurley, T. D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin, *J Mol Biol 319*, 463-477.
47. Ramakrishnan, B., and Qasba, P. K. (2002) Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity, *J Biol Chem 277*, 20833-20839.

48. Ramakrishnan, B., Ramasamy, V., and Qasba, P. K. (2006) Structural snapshots of beta-1,4-galactosyltransferase-I along the kinetic pathway, *J Mol Biol 357*, 1619-1633.
49. Bulter, T., and Elling, L. (1999) Enzymatic synthesis of nucleotide sugars, *Glycoconj J 16*, 147-159.
50. Tom, D. H., and Andrea, T. V. (1999) Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases, *Angew Chem, Int Ed 38*, 750-770.
51. Vasella, A., Davies, G. J., and Bohm, M. (2002) Glycosidase mechanisms, *Curr Opin Chem Biol 6*, 619-629.
52. Nielsen, M. M., Suits, M. D., Yang, M., Barry, C. S., Martinez-Fleites, C., Tailford, L. E., Flint, J. E., Dumon, C., Davis, B. G., Gilbert, H. J., and Davies, G. J. (2011) Substrate and metal ion promiscuity in mannosylglycerate synthase, *J Biol Chem 286*, 15155-15164.
53. Diederichs, K., and Karplus, P. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat Struct Biol 4*, 269-275.
54. Matthews, B. (1968) *J Mol Biol 33*, 491-497.
55. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J. M. (2000) X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily, *EMBO J 19*, 5269-5280.
56. Pedersen, L. C., Darden, T. A., and Negishi, M. (2002) Crystal structure of beta 1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA, *J Biol Chem 277*, 21869-21873.

2.7 ACKNOWLEDGEMENTS

The authors thank the European Synchrotron Radiation Facility (Grenoble, France) for support with the data collections. This work was also carried out with the support of the Diamond Light Source staff, funded by the Seventh Framework Programme grant nº 226716, and FCT grants PTDC/QUI/71142/2006, SFRH/BPD/29708/2006 and SFRH/BD/23222/2005. Helpful suggestions from Clemens Vonrhein (GlobalPhasing.com) and Kevin Cowtan (University of York) regarding the SHARP and BUCCANEER calculations are also gratefully acknowledged.

CHAPTER 3

Structural analysis of *Thermus thermophilus* HB27 mannosyl-3-phosphoglycerate phosphatase∗

Evidence for a concerted D_NA_N *mechanism in phosphoryl-transfer catalysis of MpgP*

Susana Gonçalves, Ana M. Esteves, Helena Santos, Nuno Borges and Pedro M. Matias †

From the Instituto de Tecnologia Química e Biologica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

∗ This chapter is adapted from two publications: [Goncalves, S., Esteves, A.M., Borges, N., Santos, H. and Matias, P.M (2011) Crystallization and preliminary X-ray analysis of mannosyl-3-phosphoglycerate phosphatase from *Thermus thermophilus* HB27. *Acta Cryst F* 67, 390-396], and [Goncalves, S., Esteves, A. M., Santos, H., Borges, N. and Matias, P.M. (2011) Three-dimensional structure of mannosyl-3-phosphoglycerate phosphatase from *Thermus thermophilus* HB27: a new member of the haloalcanoic acid dehalogenase superfamily. *Biochemistry* 50, 9551-9567].

† Author´s contributions:

The cloning and expression of mannosyl-3-phosphoglycerate phosphatase (*T.thermophilus* MpgP) from *Thermus thermophilus HB27* was carried out by Ana M. Esteves (A.M.E), under the supervision of Dr. Nuno Borges (N.B) and Prof. Helena Santos (H.S.). Purification of *T. thermophilus* MpgP was carried out by A.M.E. with support by Susana Gonçalves (S.G) and under the supervision of N.B and H.S. Protein stability assays, namely analytical size size-exclusion chromatography and differential scanning fluorimetry measurements were conducted by S.G. The crystallization and crystal soaking assays, X-ray data collection, structural determination and refinement were carried out by S.G., under the supervision of Dr. Pedro M. Matias.

3.1 ABSTRACT

Mannosyl-3-phosphoglycerate phosphatase (MpgP) is a key mediator in the physiological response to thermal and osmotic stresses, catalyzing the hydrolysis of mannosyl-3-phosphoglycerate (MPG) into the final product, α-mannosylglycerate (MG). The structure of *Thermus thermophilus* HB27 MpgP was solved by Single Wavelength Anomalous Dispersion (SAD) at 2.5 Å resolution, using the anomalous diffraction power of gadolinium, soaked into native crystals of MpgP. MpgP is a metal-dependent Haloalkanoic Acid Dehalogenase-like (HAD-like) phosphatase, belonging to the Mannosyl-3-Phosphoglycerate Phosphatase family of the Cof-type phosphatases (HAD-IIB-MPGP family; SCOP [117505]). The domains of the "open" apo- and the "closed" holo- forms of MpgP are related to each other by a hinge rotation. The dynamics of the catalytic machinery were followed using cryo-trapped reaction species to provide "crystallographic snapshots" along the reaction cycle. Enzyme activation entails a structural rearrangement of Motifs I and IV with concomitant binding of the cocatalytic Mg^{2+} ion. The closure motion of the C2B domain is subsequently triggered by the anchoring of the phosphoryl group to the cocatalytic metal center, and by Arg167 fixing the mannosyl-moiety inside the catalytic pocket. Results suggest that phosphoryl-transfer by MpgP from *T. thermophilus* HB27 involves a concerted D_NA_N mechanism (Fig 1.5) with Asp8 acting as a catalytic acid in the formation of a short-lived metaphosphate intermediate that is immediately subject to nucleophilic attack by water. These structures identify the principle mechanistic features of phosphoryl monoester transfer catalysis in members of the HAD superfamily. More generally, they suggest a possible continuum of phosphoryl transfer mechanisms, ranging from those that are purely associative to those that are purely dissociative.

3.2 INTRODUCTION

Mannosyl-3-phosphoglycerate phosphatase (MpgP; EC 3.1.3.70) is the enzyme involved in the second step of the two-step MG biosynthetic pathway and converts the phosphorylated intermediate mannosyl-3-phosphoglycerate (MPG) into MG as the final product (*1*). This enzyme shares sequence signature features with the Mannosyl-3-Phosphoglycerate Phosphatase family members (MPGP family; JCVI-CMR Accession: TIGR01486) of the aspartate-nucleophile Haloalkanoid Acid Dehalogenase-like superfamily (HAD-like; SCOP v1.75 Accession: [56783]). The HAD-like superfamily includes Mg^{2+} dependent hydrolases that catalyze the carbon- or phosphoryl-group transfer from a wide variety of substrates (*2*). Members of this superfamily are topologically characterized by a three-layer α/β sandwich having a central parallel β-sheet with strand order 54123, and by two consecutive structural motifs downstream of strand β1; a π-type helical turn known as “squiggle”, and a β-hairpin designated as “flap”. In addition, four conserved functional motifs are spatially arranged around the active site cleft in the HAD-core region, shaping an electrostatic environment that stabilizes negatively charged transition-state (TS) species during the nucleophilic substitution (*2-5*). Cap modules are the result of independent structural insertions at specific regions of the HAD-core unit and comprise three main types: the four helical bundle C1 cap; the α+β fold C2 cap, and the C0 cap corresponding to the single HAD-core domain. These are further classified into classes and subclasses, which increase in complexity as a function of the diversity of substrates and reaction systems (*2*).

The MPGP family falls within the predicted Cof-type phosphatase clade, and is characterized by having a C2 type B cap module inserted within the cross-linker region between strands β3 and β4 of the basal HAD-core domain (HAD-IIB-MPGP family; SCOP [117505]). Members of the Cof-type class are known to use sugar-phosphates as substrates, and many may also be involved in the regulation of stress response mechanisms, such as the Cof-

type stress response YhaX from *Bacillus subtilis* (*6*), and the Cof-related trehalose-6-phosphate phosphatase from *Thermoplasma acidophilum* (*7*).To date, the MPGP hydrolase family is structurally represented only by the MpgP from *Pyrococcus horikoshii* in its ""closed" " conformation (*8*) (PDB 1wzc) and by the "open" conformation of the MpgP-related YedP from *E. coli* of unknown function (PDB 1xvi; Kim *et al.,* unpublished), but known to interact with inner membrane protein YedI (*9*).

Remarkably, the catalytic core for phosphoryl-monoester transfer reaction is preserved amongst the HAD members, regardless of the nature of the substrate. Still, the detailed knowledge of the reaction coordinate landscape in these metal-dependent phosphate monoester hydrolases is currently under debate (*4, 10-12*). Whereas the structural data suggests an associative mechanism, comprising a two-step addition-elimination (IUPAC A_N + D_N) with coupled acid/base catalysis (*5*), non-enzymatic phosphoryl-monoester hydrolysis supports a dissociative S_N1-type mechanism (IUPAC D_N + A_N) with coupled acid/base catalysis (*11, 13*). The HAD-core Motifs I-IV effectively modulate a trigonal-bipyramidal electrostatic environment, hinting at the first model by favoring an equatorial expansion/polarization at the non-bridging oxygen atoms of the phosphorane intermediate (*3-5*). However, conflicting evidence arguing in favor of a dissociative mechanism arises from the observation of a trapped PO_3^- intermediate in the crystal structure of fructose-1,6-bisphosphatase (*14*), and from the measured kinetic isotope effects in other monoester phosphatases, which support the formation of a PO_3^- resonance species in the first reaction step (*9, 15, 16*). The latter studies suggested a high bond breaking order of the leaving group (β_{lg}) with a low bond order to the attacking nucleophile (β_{nuc}), and thus were explained on the basis of a concerted mechanism with higher dissociative character (*16*). An integrated perspective of these two mechanisms was given by the theoretical calculations done with representative members of the HAD-like C1 and C2 classes(*12, 17*). The transfer of the PO_3^- species is concerted and should occur with a geometric associative but an electronic dissociative character, coupled

with a proton transfer. Thus, the mechanism of phoshoryl-transfer catalysis depends on the nature of both the electrophile and the nucleophile, as well as the enzyme's ability to stabilize the transition state species, and a continuum of possible mechanisms between those purely associative and dissociative may occur (*11, 16*).

Herein we describe the crystal structure of *Thermus thermophilus* HB27 MpgP, an aspartate nucleophile HAD-like hydrolase, which belongs to the MPGP family. This work completes the structural characterization of the enzymes involved in the synthesis of MG in this organism. Crystal soaks with substrates, substrate analogues and inhibitors allowed us to probe and rationalize the relative positions of the Cap and HAD-core regions as crystallographic snapshots of the phosphoryl-transfer. In MpgP, the phosphoryl-transfer catalysis occurs by a concerted D_NA_N mechanism, assisted by a coupled proton transfer from Asp8 to the ester oxygen. Furthermore, the functional role of the conserved salt bridge between Asp8 and Lys41 in promoting the proton flux during the P—O fission is discussed in light of a versatile catalytic HAD-core which is able to synchronously modulate the hydrogen bonding network and the electrostatic potential for optimal positioning of the PO_3^- intermediate with respect to the attacking nucleophile.

The results presented may assist in resolving prior inconsistencies, as well as provide a picture of the main mechanistic aspects of phosphoryl-monoester transfer catalysis, common to other members of the HAD superfamily.

3.3 MATERIALS AND METHODS

Cloning of the Thermus thermophilus HB27 mpgP gene

T. thermophilus HB27 genomic DNA was used as the template to amplify the *mpgP* gene (GeneBank® NC_005835.1) by polymerase chain reaction (PCR). The gene was then cloned into the pKK223-3 vector (PL-Pharmacia; GeneBank® M77749.1) in-between the *EcoRI* and *PstI* restriction sites. Two primers (Thermo Scientific) were constructed and used for PCR amplification with *Pfu* polymerase (Fermentas). The sense primer was mpgP-F-5`-GGC**GAATTC**ATGATCGTCTTCACCGACCTGGA-3`, and carried the *EcoRI* restriction site, whereas the antisense primer was mpgP-R-5'-CCG**CTGCAG**TCAGGGCCCGCTCCCTCCTCGTCGG-3', and harbored the *PstI* restriction site. The PCR amplification was carried out in a iCycler thermocycler (Bio-Rad), with a 25 *μ*L reaction mixture containing 200 *n*g of genomic DNA, 0.5 *μ*M of primers, 0.25 *m*M of dNTPs, 1.5 *m*M of Mg_2SO_4 cofactor, *Pfu* buffer diluted to 1× from a 10× stock solution (Fermentas) and 1.25 U of *Pfu* enzyme (Fermentas). This mixture was pre-incubated at 368 K for 5 minutes, followed by 30 cycles of sequential steps of denaturation (368 K for 1 minute), annealing (338 K for 1 minute) and primer extension (345 K for 1.5 minutes). In the final cycle the extension step was prolonged for 7 minutes. The PCR product was further digested with *EcoRI* and *PstI* restriction enzymes and purified with the PCR DNA and Gel Band Purification Kit (GE Healthcare). The pKK233-3_mpgP vector construct was built with T4 DNA ligase (Promega) by a ligation reaction (10 *μ*L total volume) between the DNA insert and the linearized pKK223-3 vector at the corresponding *EcoRI* and *PstI* restriction sites, in a 1:4 ratio of vector to insert, respectively. Positive recombinant colonies of *E. coli* DH5α cells carrying the vector construct were selected from LB/agar plates containing ampicillin (100 *μ*g/*m*L), and the plasmid was extracted with the PlasmidPrep Mini Spin Kit (GE Healthcare). Confirmation of the inserted gene was obtained from double-strand sequencing (STAB VIDA EUROMICSsequencing service, Caparica, Portugal; http://www.stabvida.com*).*

Expression and purification of Thermus thermophilus *HB27 MpgP*

Transformation, expression and purification of MgpP was performed as previously described (*1*), with some modifications. *E. coli* BL21(DE3) strains (Novagen) bearing the plasmid construction pKK223-3_mpgP were grown in an orbital shaker at 200 rpm and at 310 K in YT medium supplemented with ampicillin (100 μg/*m*L) to an OD_{600} of 0.6-0.7, followed by induction with 1 *m*M IPTG (isopropyl-β-D-thiogalactopyranoside) for 6 hours at 310K. The cells were harvested by centrifugation (7000g, 15 min, 277 K) and resuspended in buffer A (20 *m*M Tris-HCl pH 7.6, 5 *m*M DTT, 1 *m*M EDTA). Cells were disrupted in a French Press followed by centrifugation (18000g, 40 min, 277 K) for removal of cell debris. The resulting cell extract was diluted (with 20 *m*M Tris-HCl pH 7.6, 1 *m*M EDTA) to about 10 *m*g/*m*L and heated for 20 min at 348 K to precipitate thermo-labile contaminant proteins from *E. coli*. After centrifugation (25000g, 45 min, 277 K), the supernatant was loaded onto a 70 *m*L Q-Sepharose column (GE Healthcare) equilibrated with buffer A. Elution was carried out with a linear gradient from buffer A to buffer B (20 *m*M Tris-HCl pH 7.6, 1 M NaCl, 1 *m*M EDTA, 5 *m*M DTT). Fractions containing MpgP eluted between 700 and 800 *m*M NaCl. These were pooled and dialyzed against buffer A before loading them onto a 6 ml Resource Q column (GE Healthcare) equilibrated with buffer A. Again, MpgP eluted between 700 and 800 mM NaCl from the linear gradient from buffer A to buffer B. The purest fractions were pooled and dialyzed against buffer C (20 *m*M MES-NaOH pH 6.3, 5 *m*M DTT, 1 *m*M EDTA) prior to loading them onto a 1 ml Mono S column (GE Healthcare) equilibrated with buffer C. A linear gradient was applied to buffer D (20 *m*M MES-NaOH pH 6.35, 5 *m*M DTT, 1 *m*M EDTA, 1 M NaCl), from which MpgP eluted as a single peak at about 760 *m*M NaCl. The eluted protein was concentrated to 10 mg/ml in a modified buffer D solution (20 *m*M MES-NaOH pH 6.35, 5 *m*M DTT, 1 *m*M EDTA, and 760 *m*M NaCl) and used for subsequent crystallization and gel filtration assays. Protein purity was assessed by SDS-PAGE analysis (Fig. 3.1A), which showed a single band with an apparent molecular mass of 28.2 kDa and only a small percentage of contaminant. The

MpgP activity was detected by visualizing the formation of MG derived from the hydrolysis of MPG by thin-layer chromatography as described previously (*1*).The protein concentration was determined by the Bradford assay (*18*). The molecular mass of MpgP was estimated by size-exclusion chromatography (SEC) using a 2.4 *m*L Superdex 200 3.2/30 precision column (GE Healthcare) equilibrated with buffer E (20 mM MES-NaOH pH 6.3, 760 *m*M NaCl, 1 mM EDTA and 5 *m*M DTT). The standards (GE Healthcare) used and their correspondent elution volumes (Ve) were ribonuclease (13.7 kDa; Ve = 1.89 *m*L), ovalbumin (43 kDa; Ve = 1.65 *m*L), albumin (66 kDa; Ve = 1.56 *m*L, aldolase (158 kDa; Ve = 1.43 *m*L) and ferritin (440 kDa; Ve = 1.23 *m*L). Blue dextran 2000 (GE Healthcare) was used to determine the void volume of the column (Ve = 1.01 *m*L). All of the assays were performed at a constant flow rate of 0.1 *m*L/min. The MpgP protein eluted at Ve = 1.59 *m*L, between the elution profiles for albumin and ovalbumin, suggesting that a homodimeric structure is the prevalent oligomeric state under the running conditions (Fig. 3.1B).

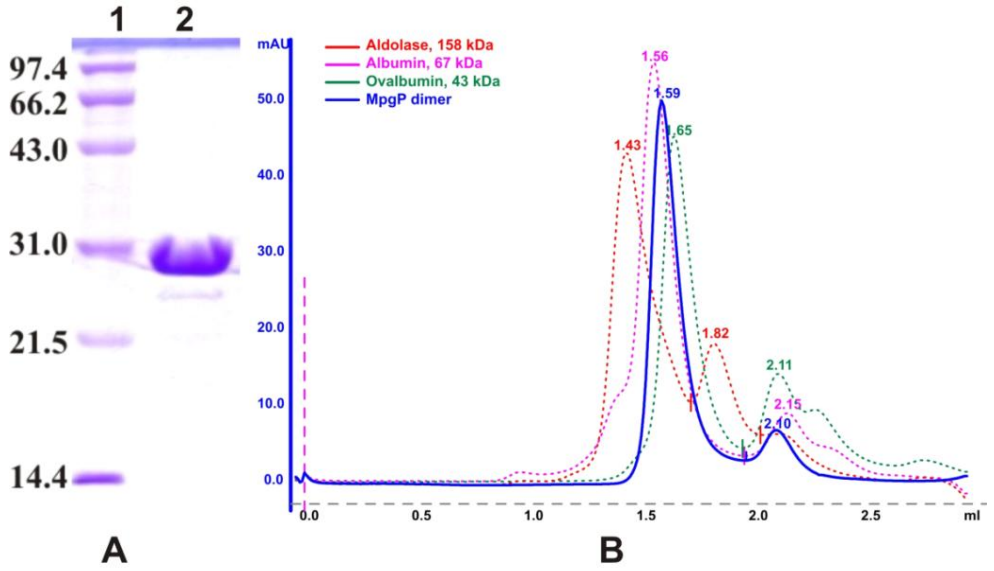


Figure 3.1 A) **15% SDS–PAGE of pure recombinant *T. thermophilus* HB27 MpgP**. Lane 1, Low Range Molecular Mass Markers (Bio-Rad) consisting of the protein markers phosphorylase b (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (43 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa) and lysozyme (14.4 kDa). Lane 2, MpgP monomer migrating according to its molecular mass (28.18 kDa). **B) Elution profile of MpgP loaded onto an analytical 2.4 *m*L Superdex 200 3.2/30 precision column**. The MpgP elution profile is represented as a full curve in blue. Elution profiles of protein standards are shown as dashed curves for comparison: aldolase (158 kDa) in red, albumin (67 kDa,) in pink and ovalbumin (43 kDa) in green. In addition to their standard elution peaks, degradation products are also eluted. mAU, milliunits of absorption.

Differential Scanning Fluorimetry (DSF) and thermal shift assay

To address the solubility and stability of MpgP as a function of salt, pH and buffer composition, DSF analysis (*19*) was performed in a protein sample eluted from the MonoS column (1.2 *m*g/*m*L) containing the probe dye SYPRO Orange (Invitrogen) diluted to 50x from the original stock solution (5000x) and distributed in a Low 96-well White Multiplate PCR Plate (Bio-Rad). In each well, the protein-sample mixture (2 *u*L) was diluted 10x (to a final volume of 20 *u*L) by adding 18 *u*L of the corresponding buffer solution from a home-made 96-condition sparse-matrix screening kit. The buffer screening kit was prepared from the commercial JBS solubility kit buffers (Jena Bioscience). For each condition, NaCl was included at concentrations of 0, 150 and 500 *m*M. The last condition contained 150 *m*M KCl instead of NaCl. The plate was sealed with Microseal type B film seal (Bio-Rad) and submitted to a temperature gradient scan (293-363 K) in a iQ5 Real Time PCR machine (Bio-Rad). The fluorescence emission from the reporter dye was measured every 2 K of increasing temperature using the Cy3 fluorophore. The scanning protocol consisted of temperature stabilization at 293 K for 1 min in the first cycle, followed by a second cycle of 71 steps up to 363 K; in each step, the

temperature gradient was held for 10 s before data acquisition. The third and last cycle consisted of lowering the temperature back down to 293 K to cool down the system before sample removal. The melting curves (Fig. 3.3) were analyzed with the iQ5 Optical System Software v.2.0 (Bio-Rad) and melting temperature values (Tm) were calculated with the Differential Scanning Fluorimetry Analysis tool (DSF v.2.5 for Bio-Rad IQ5; ftp://ftp.sgc.ox.ac.uk/pub/biophysics; (*19*).

Effect of NaCl, EDTA and DTT in MpgP oligomerization and stability

MpgP protein eluted from the MonoS column was concentrated to 8.6 *m*g/*m*L in modified buffer D solution (see above) and was used for analytical SEC studies performed with a Superdex S75 3.2/30 precision column (GE Healthcare). Prior to column injection, the protein was diluted 5x (10 *u*L total volume) in the corresponding gel-filtration buffer solution used during each assay. The standard molecular-mass curve was determined using the globular protein standards (GE Healthcare) ribonuclease (13.7 kDa; Ve = 1.52 *m*L), chymotrypsinogen (25 kDa; Ve = 1.44 *m*L), ovalbumin (43 kDa; Ve = 1.26 *m*L) and albumin (66 kDa; Ve = 1.18 *m*L) in 20 *m*M MES-NaOH pH 6.3, 760 *m*M NaCl and 1 *m*M EDTA (Fig. 3.4A). Blue dextran 2000 (GE Healthcare) was used to determine the void volume of the column (Ve = 1.03 *m*L). The effect of NaCl, DTT and EDTA on the Ve of the protein standards was first assessed and was found to be negligible in comparison with the values obtained when using the manufacturer's recommended buffer (20 *m*M Na_2HPO_4/ KH_2PO_4 pH 7.2, 150 *m*M NaCl). All of the assays were performed at a constant flow rate of 0.1 *m*L/min in the MES-NaOH pH 6.3 buffer system. The variation of the elution profiles with buffer composition is shown in Fig. 3.4B.

MpgP crystallization and cryo-conditions

Preliminary crystallization trials were done at the nanoscale with the commercially available Morpheus™ crystallization kits (Molecular Dimensions) (*20*), using a Cartesian Crystallization Robot Dispensing System (Genomics

Solutions) and round-bottom Greiner 96-well CrystalQuick™ plates (Greiner Bio-One). Three crystallization drops per condition screened were prepared by mixing 100 *n*L reservoir solution with 100 *n*L protein solution. MpgP at 10 *m*g/*m*L in its concentration buffer was tested, as well as with added 5 *m*M $MgCl_2$, 5 *m*M sodium/potassium phosphate pH 6.8 and with added 5 *m*M $MgCl_2$, 5 *m*M sodium/potassium phosphate pH 6.8, 1 *m*M tris-(2-carboxyethyl)phosphine (TCEP). The drops were equilibrated against 100 μL reservoir solution. Promising plate-shaped and prism-like crystal clusters were found for all three of the MpgP protein samples in distinct conditions from the 96-well sparse matrix.

In the scale-up trials a crystallization solution with the following composition was used (*20*): 0.02 M of each of the carboxylic acids sodium formate, ammonium acetate, trisodium citrate, sodium potassium L-tartrate and sodium oxamate as additives, 0.1 M MES-imidazole pH 6.5 (in a 1:1 molar ratio) as a buffering system and 30% of a precipitant mixture composed of 20% ethyleneglycol (EDO) and 10% polyethylene glycol 8000 (PEG 8K). Native protein at 10 *m*g/mL with added 5 *m*M $MgCl_2$, 5 *m*M sodium/potassium phosphate and 2 *m*M DTT was mixed with the crystallization solution in a 1:1 ratio of protein to reservoir solution and 2 μL drops were set up at 293 K in 24-well crystallization plates using the sitting-drop vapour-diffusion technique. The drops were equilibrated against 500 μl reservoir solution. Prism-like crystals (Fig. 3.2) developed within 5 days, reaching dimensions of 30x 50 x 400 μm. Since the crystallization solution is already cryoprotecting (*20*), native MpgP crystals were harvested and immediately flash-cooled in liquid nitrogen prior to data collection. The crystallized protein was confirmed as MpgP by N-terminal sequencing (Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Oeiras, Portugal) of a dissolved crystal from the same drop as that used for X-ray diffraction data measurements. For crystals soaked with gadolinium (III) chloride hydrate ($GdCl_3.xH_2O$), 2 *m*M of the lanthanide solution were used instead of $MgCl_2$ during co-crystallization.

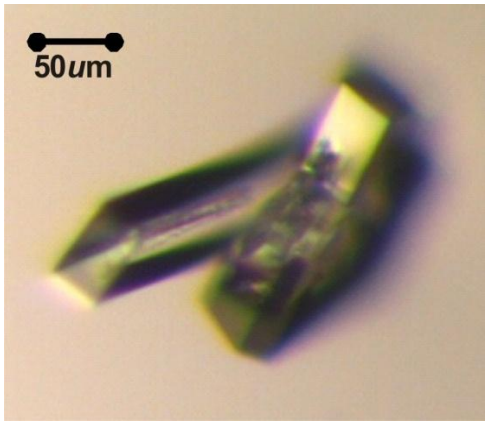


Figure 3.2 Native MpgP crystals used for data collection

Crystal soaking procedure

The MpgP crystal used for phasing, *i.e*, the [Gd^{3+}] dataset, was gradually transferred through six different stabilizing solutions with chemical composition similar to that used for protein crystallization, but where the carboxylic acids (CarAc) mixture was gradually replaced by $GdCl_3.xH_2O$ (Table 3.1). The removal of the CarAc mixture was done as an optimization strategy to increase the yield of bound ligand, by preventing the adventitious metal chelation by some of these acids. The MES-NaOH buffer pH 6.5 was used instead of the MES/Imidazole pH 6.5 buffer and the final soaking solution contained 35% (v/v) precipitant mixture (EDO-PEG 8K) to ensure proper cryo-protection. The final soaking with Gd^{3+} was performed at 293 K over a period of 14 hours (Table 3.2), and the crystals were subsequently flash-cooled in N_2(l) prior to data collection.

The crystal soaks with substrates and with or without the presence of inhibitors followed a similar procedure to that described above, but these ligands were added only in the final two crystal transfer solutions. As control experiments, one *pseudo*-Michaelis ternary complex and another mirroring the pentavalent TS species of the reaction coordinate were prepared by soaking crystals with MG and either orthophosphate or orthovanadate analogues,

respectively. The composition of the soaking solutions and the final ligand concentrations, as well as the duration of the soaks, is described in Table 3.2 for each dataset. Hereafter, each dataset will be designated accordingly to the soaking conditions used. All the soaks were done at 293 K. Stock solutions used for soaking were respectively: 0.5 M $GdCl_3.xH_2O$ (Sigma-Aldrich) in MES-NaOH pH 6.5; 0.5 M $VOSO_4.5H_2O$ (Merck) in water; 0.2 M Na_3VO_4 (Sigma-Aldrich) at pH 10.0 in water; 0.8 M $MgCl_2$ (Sigma-Aldrich) in water; 0.1 M α-D-mannosylglycerate (MG) at 0.1 M in water; 0.019 M α-D-mannosyl-3-phosphoglycerate (MPG) in water. The MPG substrate was produced with *T. thermophilus* HB27 MpgS enzyme purified as previously described in Chapter 2. The experimental conditions applied for reaction and isolation of MPG followed those established by Empadinhas and co-workers (*1*), with slight modifications; MPG eluted fractions from the QAE-Sephadex A25 column were detected by Thin Liquid Chromatography (TLC), and analyzed by ^{1}H NMR. The MG substrate was obtained as previously described by Borges and co-workers (*21*).

Data collection and processing

A highly redundant Single Wavelength Anomalous Dispersion (SAD) data set to 2.5 Å resolution was measured *in-house* with Cu K_α radiation (λ=1.5418 Å) at 100 K from a crystal soaked with Gd^{3+} ([Gd^{3+}] dataset), using a Bruker-AXS Proteum Pt135 CCD detector system coupled to a Bruker-AXS Microstar-I rotating anode X-ray generator with Montel mirrors. The data were integrated with SAINT and scaled with SADABS as part of the Bruker-AXS Proteum Software Suite. Diffraction data statistics were obtained with XPREP (Bruker-AXS). A summary of the data collection and processing statistics is given in Table 3.3. The remaining datasets were all collected at 100 K at the ESRF (Grenoble, France) ID14-4 and ID23-1 beamlines, using an ADSC Quantum Q315r detector except for the [Mg^{2+}-MG-VO_4^{3-}] dataset, which was collected at DIAMOND (Didcot, U.K.) I03 beamline, using a Pilatus 6M detector. Diffraction images were processed with the XDS Program Package

(*22*), and the data collection statistics are summarized in Table 3.4. The data were further processed for use with the CCP4 Program Suite (*23*). Matthews coefficient calculations (*24*) suggested the presence of 2 molecules in the asymmetric unit of all crystal structures, with values close to those calculated for the [Gd^{3+}] dataset (V_m=2.31 $Å^3Da^{-1}$ and a predicted solvent content of 46.7 %).

Structure solution and crystallographic model building

Using the HKL2MAP (*25*) graphical user interface, the [Gd^{3+}] dataset was scaled and analysed with SHELXC (Sheldrick G., personal communication), the Gd^{3+} heavy atom substructure was determined with SHELXD (*26*) and the phase problem solved with SHELXE (*27*). The best solution from SHELXD in 100 trials gave 6 Gd^{3+} sites with a correlation coefficient of 46.9%, four of them with a relative occupancy about half that of the top two. The SHELXE calculations gave a clear discrimination between the correct and the inverted substructure solutions, and the phases derived from the SAD data were improved using the maximum-likelihood heavy-atom parameter refinement in autoSHARP (*28*). The autoSHARP calculations found 5 additional sites, and a subsequent optimizing density modification procedure using SOLOMON (*29*) suggested a solvent content of 40.5% and two monomers in the asymmetric unit. Centroid SHARP phases were further improved by density modification with DM (*30*) with 2-fold non-crystallographic symmetry averaging (NCS). The initial NCS operator was obtained with PROFESSS from the top 6 Gd^{3+} sites. A random 5% sample of the reflection data was used for R-free calculations (*31*) during model building and refinement. Using the 2.5 Å density-modified phases from DM, 463 of the expected 518 protein residues in the asymmetric unit were built and sequenced automatically with Buccaneer/REFMAC (*32, 33*) leading to values of R and R-free of 0.305 and 0.351. The model was then corrected and completed using Coot (*34*) prior to refinement. The main statistics from the autoSHARP/SOLOMON/DM calculations are included in Table 3.3.

The structures of the apo-MpgP and of the crystals soaked with substrates and inhibitors were determined by molecular replacement with PHASER (*35*) using the refined protein chain coordinates of the [Gd^{3+}] crystal structure.

Crystallographic refinement

The structures were refined using the amplitude-based Maximum-Likelihood target function with automatic weight optimization procedure as implemented through the graphics user interface of the PHENIX v.1.6.1 software package (*36*). In general, the first cycle comprised a rigid body refinement followed by individual coordinate and isotropic atomic displacement parameter refinement. In the subsequent refinement stages, TLS (translation-libration-screw) rigid body refinement of atomic displacement parameters was carried out, followed by refinement of individual isotropic B-factors, in addition to individual atomic coordinates. In the case of the [Gd^{3+}] structure, and due to the lower data resolution, group isotropic B-factors were refined for the protein chains instead of individual isotropic atomic displacement parameters (ADPs). Two groups were defined per protein residue, one for the main-chain atoms and the other for the side-chain atoms of each protein chain residue. For each crystal structure, four TLS groups per monomer were parameterized using the TLS Motion Determination (TLSMD) server (*37*). The definition of the TLS groups comprised the segmentation of the two protein domains: the first, second and fourth groups roughly corresponded to the HAD-like domain whereas the third contained at least part of the cap C2B domain (see the subsection "Structure of MpgP" in Results for the definitions of the HAD-like and cap C2B domains). Whereas the first and fourth TLS groups had similar boundaries for all structures, the separation between the second and third TLS groups varied: for example, while in chain B of the [Gd^{3+}] crystal structure the third group closely matched the cap C2B domain, in the control [Mg^{2+}-MG-HPO_4^{2-}] structure, part of this domain was included in the second TLS group. The inclusion of NCS restraints did not improve the refinement statistics or the

model quality. Monomer library descriptions for new ligands were created with LIBCHECK (*38*) through the Sketcher interface implemented in the CCP4i Graphical User Interface (*39*). Ligand and metal coordination geometry restraints were further reviewed with REFMAC5 (*33*). During refinement, the models were periodically inspected and corrected with Coot (*34*) against SigmaA weighted $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps. Inclusion of ordered solvent molecules was done with ARP/wARP (*40*), followed by inspection in Coot (*34*). Except for the [Gd^{3+}] structure, in the final refinement, automatic refinement of water occupancy and constrained-group occupancy refinement for non-protein atoms were included in the refinement protocol. The stereochemical quality of the model was assessed with MolProbity (*41*). The models were almost complete, varying in length after residue Arg250 except for the apo-MpgP crystal structure that presented poor electron density at residues Asp13-Leu18 in chain B, which were thus not included in the model. A summary of the refinement statistics, model composition and stereochemical quality is presented in Table 3.5. Hereafter, each structure will be designated according to the soaking conditions used. Coordinates and structure factors were deposited with the Protein Data Bank in Europe (PDBe) (*42*), and the respective accession codes are: 3zty for the Gd^{3+} soaked structure; [Gd^{3+}], 3ztw for apo–MpgP, 3zu6 for the [Mg^{2+}-MPG] crystal 1, 3zup for the [Mg^{2+}-MPG] crystal 2, 3zw7 for the [Gd^{3+}-MPG] crystal 1, 3zwd for the [Gd^{3+}-MPG] crystal 2, 3zw4 for the [Mg^{2+}-DMG-PO_4^{2-}] (Control) and 3zx5 for the [Mg^{2+}-DMG-VO_4^{3-}] (Control).

3.4 RESULTS

Size Exclusion Chromatography (SEC) and Differential Scanning Fluorimetry (DSF) assays

Application of the differential scanning fluorimetry (DSF) methodology(*19*) to the screening of 96 sparse-matrix buffer conditions allowed a quick analysis of the protein stability as a function of salt, pH and buffer. The results provided useful information regarding the choice of the storage buffer for the protein. In most of the conditions tested for MpgP the observed thermal unfolding profile was represented by a single sigmoidal curve apparently corresponding to a two-state transition. The best results were obtained for conditions containing high salt (500 *m*M NaCl) and within the pH interval 5.5 - 7.0 (Fig. 3.3), corresponding to T_m values ranging between 354 and 357 K. Interestingly, when sodium citrate pH 5.5 was present in the buffer system the protein exhibited a similar unfolding profile (T_m = 355 K) regardless of the presence of salt. This suggests that the carboxylic groups in sodium citrate have a stabilizing effect on MpgP, probably competing with salt in a type of interaction analogous to MonoS elution. Interestingly, the MpgP that eluted from the MonoS column was already in a buffer composition resembling those giving highest stability as found from the DSF assay and was also able to crystallize readily.

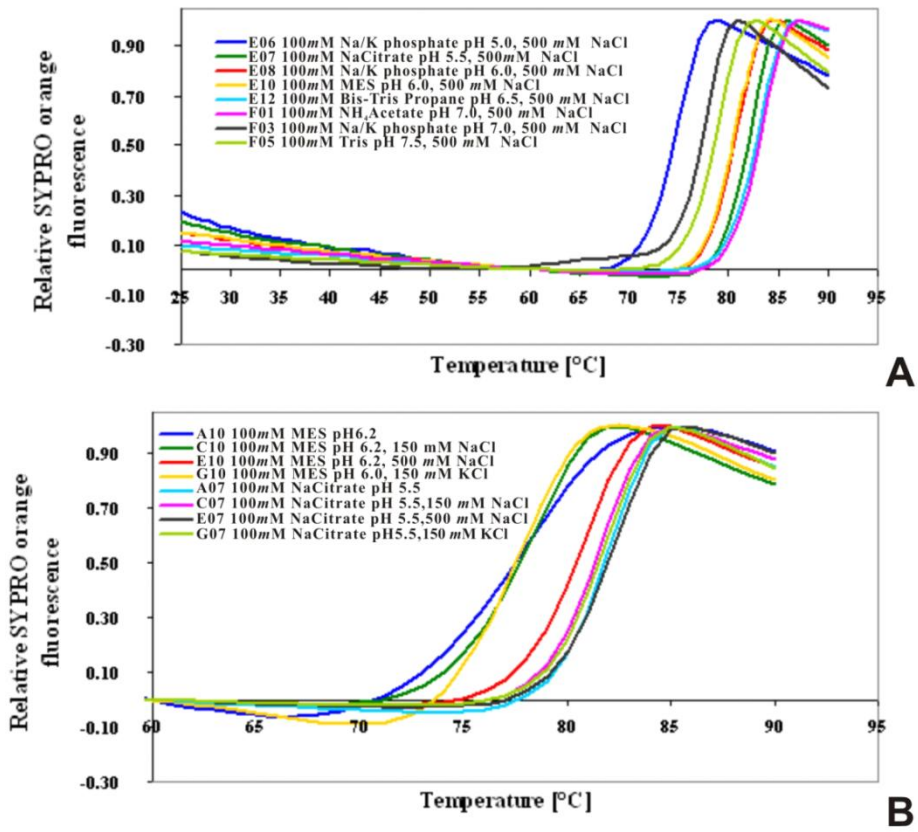


Figure 3.3 DSF thermal denaturation curves for MpgP. The buffer composition does not take into account the 10% dilution effect during protein-sample dilution (see "Materials and methods"). **A)** MpgP stability as a function of pH at high ionic strength. The thermal unfolding profile is shown only within the pH interval 5.0 - 7.0 as the protein was less stable at other pH values. The chemical nature of the buffer also had an effect on the stability of MpgP as observed for the profiles between curves F01 and F03. **B)** MpgP stability as function of salt. In the MES buffer system the stability of MpgP increases significantly with an increase in ionic strength (curves A10, C10 and E10), while the substitution of equivalent amounts of NaCl by KCl have little effect (curves C10 and G10). Sodium citrate pH 5.5 buffer (curves A07, C07, E07 and G07) was an exception since MpgP exhibited similar unfolding profiles regardless of the presence of salt.

The role of NaCl, EDTA and DTT in MpgP dimerization and solubility

At the beginning of the optimization process of MpgP purification, the protein showed a tendency to aggregate and could not be concentrated above 1.2 *m*g/*m*L. This issue was overcome by direct elution from the MonoS column at about 760 *m*M NaCl, 5 *m*M DTT and 1 *m*M EDTA, thus allowing the protein to be concentrated to at least 10 *m*g/*m*L. As indicated by its elution profile from the analytical Superdex 200 3.2/30 precision column, this soluble form is probably a dimer in this buffer system, with an estimated molecular mass MM_{ca}l of 65 $\pm$ 9 kDa (Fig. 2b). The effects of NaCl, DTT and EDTA were then assessed by size-exclusion chromatography (SEC) using an analytical Superdex 75 3.2/30 precision column (GE Healthcare) in the MES-NaOH pH 6.3 buffer system (Fig. 4). The elution profile of MpgP varied greatly with the ionic strength and with the presence of EDTA; the MpgP dimer (Ve = 1.24 *m*L) only eluted at 760 *m*M NaCl, 1 *m*M EDTA, while in the absence of either one of these two components low-molecular-mass products eluted. This suggests that strong electrostatic interactions were established between the protein (either folded or unfolded) and the column matrix. This is not surprising, since the theoretical isoelectric point of the MpgP monomer is 7.8 and, as confirmed by its strong binding affinity to the MonoS column, in MES-NaOH pH 6.3 buffer the dimer is very likely to have a total net positive charge in solution which is sufficiently high to establish this type of interaction. The addition of DTT also improved the solubility of the protein by preventing the formation of nonspecific disulfide bridges. However, after heating a protein sample to the optimal temperature for catalysis [348 K;(*1*)] the resulting elution profile suggested dimer destabilization together with protein degradation (data not shown). The nonconserved cysteine residue (Cys144) found in a variable region of the *T. thermophilus* MpgP sequence may therefore play a role in the dimerization of MpgP. Our results suggest that *T. thermophilus* MpgP requires a high ionic strength to become properly folded and is most stable in solution as dimers, the formation of which possibly involves a disulfide bridge.

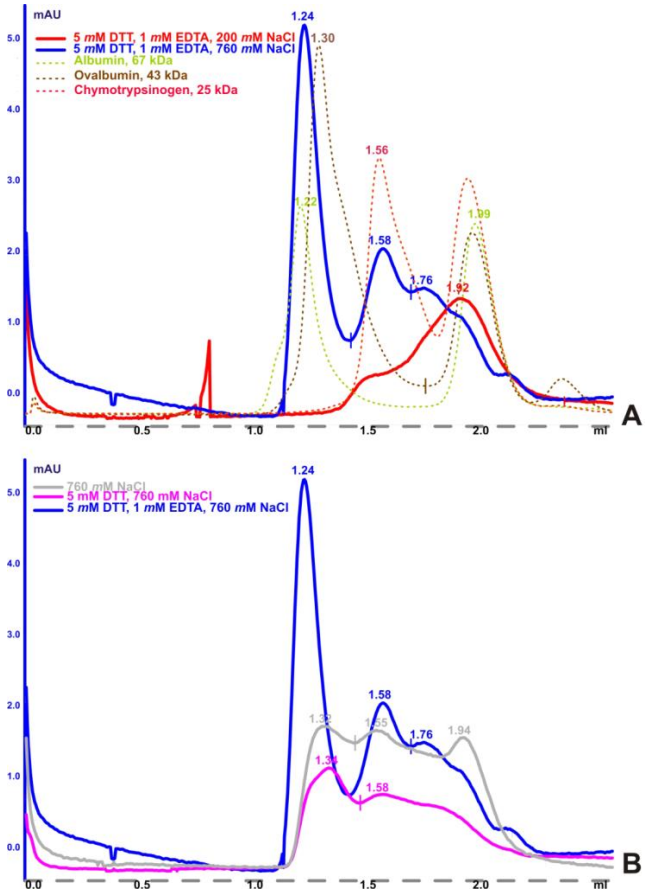


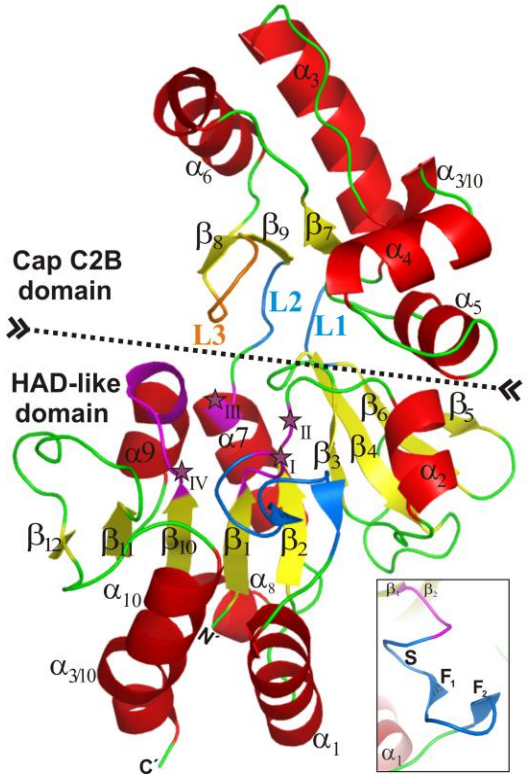
Figure 3.4 MpgP elution profiles from the SEC analysis with Superdex 75 3.2/30 precision column. MpgP elution profile represented as full curves. The upper left text in each panel indicates the gel-filtration buffer composition in which the MpgP was eluted and is color-coded according to the colors of the MpgP elution curves. The MES-NaOH pH 6.3 buffer system was used in all the gel-filtration runs. **A -** Effect of NaCl in the elution of MpgP. The elution profiles of protein standards are shown as dashed curves for comparison. In addition to their standard elution peaks, lower molecular-mass degradation products are also eluted. **B -** Effect of EDTA and DTT on the MpgP elution profile. *m*AU stands for mili-units of absorption.

Structure of MpgP

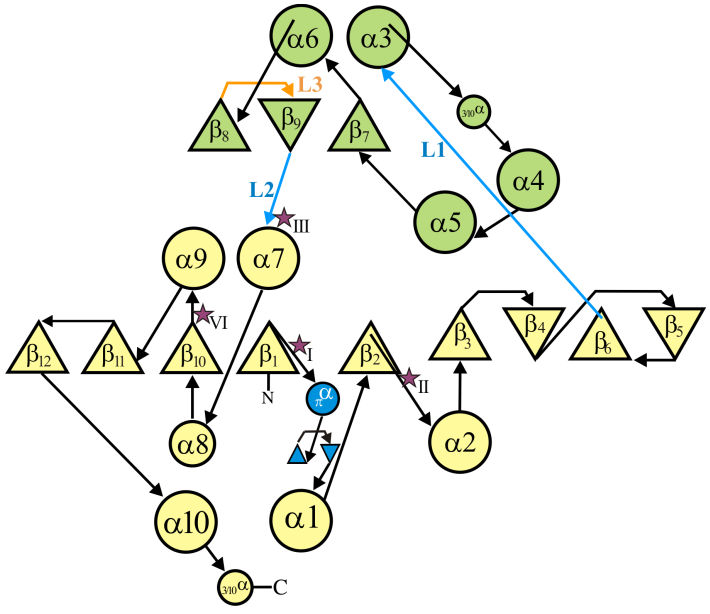
MpgP crystallized as a homodimer, with each monomer related by a 2-fold non-crystallographic symmetry (NCS) axis, in both monoclinic and orthorhombic crystal structures. In the monoclinic crystal structures, the 2-fold NCS is oriented perpendicularly to the unique *b*-axis, and the homodimers form a right-handed helical assembly along the crystallographic *c*-axis. This homodimer arrangement is not present in the orthorhombic crystal structure. A secondary structure matching against all PDB, carried out with DALI (*43*) and SSM (*44*), and using the coordinates of chain A from the apo-MpgP and chain B from the control [Mg^{2+}-MG-HPO_4^{2-}] structures, confirmed the expected close relationship to the members of the predicted hydrolase Cof family (SCOP v1.75 Accession: [82388]) of the Haloalkanoid Acid Dehalogenase Superfamily (HAD-SF). Proteins from this family share a C2B cap module characterized by an "open" -face $\alpha+\beta$ unit with at least a three-stranded anti-parallel β-sheet flanked by two oppositely oriented α-helices (*45*). The most similar protein structures identified with both search engines were those sharing the mannosyl-3-phosphoglycerate phosphatase protein domain (putative HAD-SF-IIB-MPGP family, SCOP v1.75 Accession: [117505]), that includes the orthologous MpgP from *P. horikoshii* (PDB 2zos; 1wzc; (*8*) and the MpgP-related YedP from *E. coli* K12 (PDB 1xvi). All three proteins share the same fold, supporting the inclusion of *T. thermophilus* HB27 MpgP into this family, and suggesting divergence from a common ancestor. Other structural relatives from the predicted hydrolase Cof family included members of the phosphoglycolate phosphatase subfamily (PGPase), sucrose-phosphatase (SPP) subfamily as well as several predicted phosphoric-monoester sugar hydrolases of unknown function. The Cof hydrolases and Cof-related members are known to act with a wide spectrum of phospho-sugar compounds.

Each *T. thermophilus* MpgP monomer is composed of two domains (Fig. 3.5): the HAD-like domain (Met1-Ser84; Asp176-Pro259) and a cap module of the C2B subclass (Tyr89-Ala172), which is inserted between strand β6 and helix α7 of the HAD-like domain via two flexible loops referred to as L1

(Leu85-Trp87) and L2 (Lys173-Ala175). The architecture of the HAD-like domain comprises the basal HAD-core unit (*2*) (residues Met1-Gly61; Lys177-Val219), which is composed of a 3-layered α/β sandwich with a central β-sheet formed by 5 parallel β-strands (β11 β10 β1 β2 β3), and is extended through the insertion of additional strands, arranged in the following order: β12 β11 β10 β1 β2 β3 -β4 β6 -β5, with strands β4 and β5 antiparallel to the rest. Strand β1 is immediately followed by two consecutive highly flexible structural elements: a rare π-helical turn (residues Leu7-Leu12), and a type IV β-turn (residues Asp13-Glu17). These correspond respectively to the “squiggle” and “flap” signature motifs present in the HAD-SF members (*2*). The C2B cap module, comprises a typical α+β Cof-core region (residues Tyr89-Ala103 and Glu139-Ala172) with a three-stranded anti-parallel β-sheet (strands β8, -β9, β7) surrounded by two helices (α3 and α6), extended by the insertion of two additional oppositely oriented helices (α4, α5) which is characteristic of the MPGP family. A cavity enclosing the active site is located between the C2B cap and the HAD-like domain.



A



B

Figure 3.5 Three-dimensional structure of MpgP from *T. thermophilus* HB27. (**A**) Cartoon representation of the MpgP monomer. The dashed line indicates the interface between the HAD-like and C2B cap domains. Flexible loops L1 and L2 (cyan) connect both domains, between strand β6 and helix α7. At the C2B cap domain, loop L3 is located at the three-stranded antiparallel β-sheet (β8-β9-β7) of the Cof core unit and is involved in recognition and binding of the mannosyl moiety. The core region of the HAD-like domain consists of an α/β-sandwich with a parallel central *β*-sheet (β11β10β1β2β*3*) and two consecutive structural motifs (cyan): squiggle (S) and flap (F), immediately following strand β1. Conserved catalytic motifs I-IV are denoted with purple stars. The inset shows in detail the structural elements of the S and F motifs: a π-helical turn and a β-turn. This panel was prepared with PyMOL (http://www.pymol.org). (**B**) Topology diagram of the MpgP monomer. The HAD-like domain is colored yellow, and the C2B cap domain is colored green. Secondary structure catalytic elements correspond to those shown in panel A; α-helices are represented by circles and β-strands by triangles. The secondary structure elements were assigned on the basis of DSSP algorithm (*46*) implemented in PROCHECK3 (*23, 47*) and the PROMOTIF version 3.0 output of (*48*).The topology diagram was based on TOPS (*49*).

The stability of the 2-fold NCS homodimer was assessed using the Protein Interfaces, Surfaces and Assemblies (PISA) service (*50*), as the measured gain in solvation energy of the buried residues upon interface formation (ΔG^i), and its biological relevance was estimated by the complexation significance score (CSS). For the apo-MpgP structure, the buried area was 7% (1540 $Å^2$) of the total surface area (22920 $Å^2$) of the homodimer, and these values were similar to those obtained for the other structures. This corresponded to estimated ΔG^i values ranging between -6 and -10 kcal/mol, and to a rigid-body entropy change at dissociation ($T\Delta S^{diss}$) of 13 kcal/mol. The

interface has a hydrophilic character whereby a mean value of 16 hydrogen bonds and 18 salt bridges are formed upon dimer assembly, in addition to a disulfide bridge *via* two cysteine residues related by 2-fold NCS, Cys144. In all crystal structures the estimated free energy of dissociation (ΔG^{diss}) varied between 8 and 10 kcal/mol, with the exception of the [Mg^{2+}-MPG] crystal 1 and crystal 2 structures, which presented much lower ΔG^{diss} values of -0.7 and 4.1 kcal/mol, respectively. A CSS value of 1.0 suggests that the dimeric assembly is stable, probably corresponding to the stable dimer form in solution mentioned above. In contrast, most of the proteins structurally homologous to *T. thermophilus* MpgP are predicted to be monomeric in solution. Cys144 residue is conserved only among the *T. thermophilus* species, despite of its location in the structurally conserved 3_{10} helical turn that immediately precedes helix α6 of the basal Cof unit. The dimer stability is compromised in the *T. thermophilus* MpgP crystal structures where the Cys144 residues are fully reduced, as observed by the 10-fold decrease of the CSS (CSS=0.1) and the change in ΔG^{diss} to about -0.7 kcal/mol in the [Mg^{2+}-MPG] crystal 1 structure. However, this reduction of the cysteine residues with concomitant dissociation of the disulfide bridge may be an artifact caused by radiation damage during the synchrotron data collections (*51*).

Structural differences between apo-MpgP and holo-MpgP

The superposition between the HAD-like and the C2B cap domains of the two monomers in the monoclinic homodimer was done with CCP4 LSQKAB (*23, 52*). The apo-MpgP and the control [Mg^{2+}-MG-HPO_4^{2-}] structures were selected as representatives and are hereafter referred to as apo-MpgP and holo-MpgP respectively. A good C^{α} fit was observed, as shown by the average root mean squared (r.m.s.) coordinate deviations for apo-MpgP (0.80 Å for the HAD-like domain and 0.17 Å for the C2B cap domain), and holo-MpgP (0.73 Å for the HAD-like domain and 0.31 Å for the C2B cap domain). The largest deviations were seen for the solvent-accessible regions and especially for the flexible loops that surround the catalytic pocket. In the HAD-

like domain, these comprise the “squiggle” and “flap” motifs and a 3_{10} helical turn (Asp203-Asp205) preceding helix α9, while the C2B cap domain includes the β-turn that connects strands β8 to β9 (Gly165-Phe168), herein referred to as Loop 3. Some structural deviations were also seen at the HAD-like domain boundaries involving the two helices (α1 and α10) in the top-layer of the α/β/α sandwich, the partially disordered strand β5 and its flanking residues (Gly74-Gly79), and the long C-terminal loop (Gly220-Pro234) that brings strand β12 and helix α10 towards the core of the central β-sheet. The higher flexibility of these regions is reflected in their higher B-factors relative to those of the core domains (Fig. 3.6).

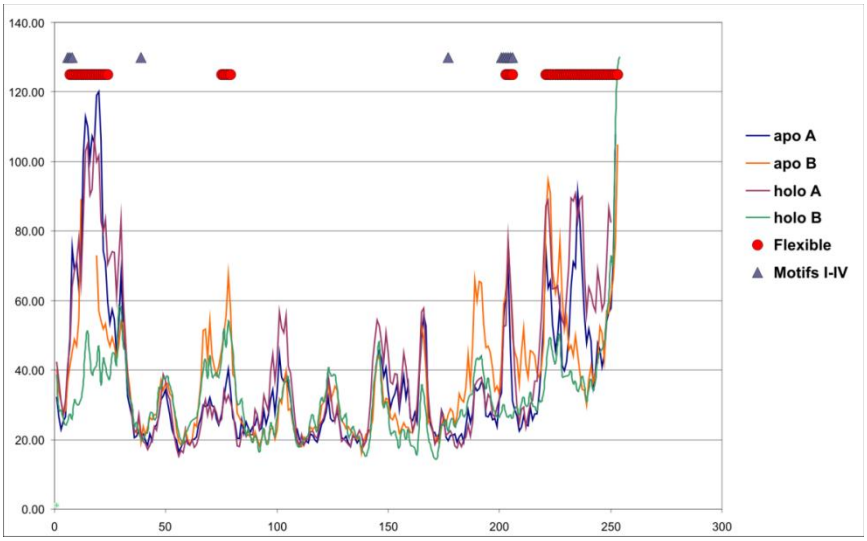


Figure 3.6 Mean main chain thermal B-factors for chains A and B of apo-MpgP and holo-MpgP (control [Mg^{2+}-MG-HPO_4^{2-}]) structures.

When superposing the corresponding domains of the apo-MpgP and the holo-MpgP structures, small deviations at their core regions were also observed. The apo- and holoenzyme structures clearly describe two distinct conformational states, whereby a rotation of the C2B cap domain leads from an “open” to a “closed” form of MpgP (Fig 3.7). The rotation of the C2B cap

domain in apo-MpgP (chain A) onto its position observed in the holo-MpgP (chain B), was analyzed with DynDom in CCP4 (*23, 53*) using default parameters. This motion was rationalized as a hinge bending of about 11º around an axis which intersects the plane of the core β-sheet between strands β3 and β4 and is oriented towards the two flexible loops L1 and L2 (Fig. 3.8A). The bending residues include those in the partially disordered strand β5 and the subsequent β-turn (Arg75-Arg81), and the mechanical hinge residues (Leu85-Ala86 and Gly174-Ala175) which are part of loops L1 and L2 respectively. However, a similar analysis of the C2B cap domain rotation in apo-MpgP (chain A) onto holo-MpgP (chain A) reveals a more complex and concerted motion involving at least three hinge axes, one of which was interpreted as a hinge motion of about 7º around an axis equivalent to that described for holo-MpgP chain B (Fig. 3.8B), and corresponds to a "*quasi*-closed" state of holo-MpgP chain A. These results suggest an internal flexibility of the C2B cap domain in MpgP. The crystal structures obtained from the remaining datasets fall within these three major conformational states: chain B is prevalently found with its bound ligand and in a "closed" conformation, whereas chain A remains in its "open" state for the shorter soaks or in a *quasi*-"closed" state for longer soaks. These differences are most likely due to structural constraints arising from the crystallographic packing in the monoclinic crystal structures. To investigate this, Solvent Accessible Surface (SAS) calculations were done with AREAIMOL (*23*) for the apo-MpgP to assess differences due to crystal contacts. An asymmetry in the distribution of accessible surface variations between monomers was observed, particularly in the cap domain. While no significant interactions are observed for chain B, the cap domain of chain A interacts with the C-terminal region of the HAD-like domain from two symmetry-equivalent molecules. This interaction includes strand β5 and its flanking residues, which contribute to the hinge bending motion, and probably restricts this motion, thus favoring a predominantly "open" conformation for holo-MpgP chain A in all the analyzed crystal structures. The analysis of these interfaces with PISA (*50*) showed the nature

of the interactions to be predominantly electrostatic. In contrast, a single interface is observed in the orthorhombic crystal structure, between two monomers (chain A) related by a unit cell translation along the crystallographic *a*-axis, which is formed by residues from the Cof-core region in one monomer and residues flanking strand β5 in the second monomer. In this structure, a nearly perfect superposition is found between one monomer and its 2-fold NCS counterpart, and both holo-forms were found in their "closed" conformation, as shown by their hinge motion relative to apo-MpgP (11º for chain A and 12º for chain B).

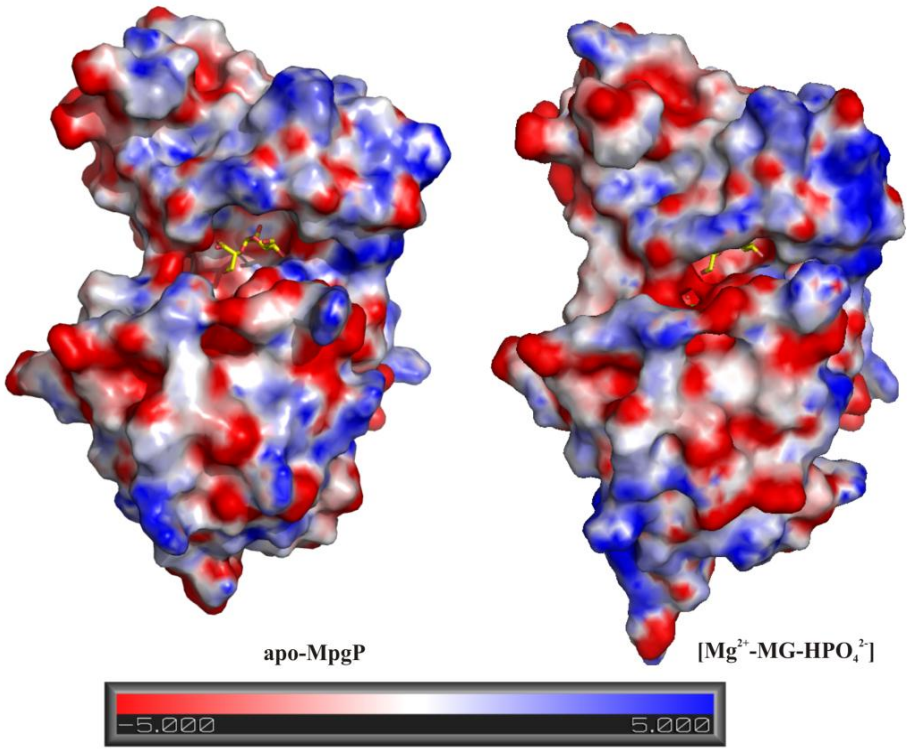


Figure 3.7 Electrostatic potential at the molecular surfaces for the apo-MpgP and holo-MpgP structures. Apo-MpgP is represented by its chain B and holo-MpgP by chain B of the control [Mg^{2+}-MG-HPO_4^{2-}] structure. The structures were superposed at their HAD-like domains and illustrate the changes that take place, particularly in the catalytic pocket region, with the hinge rotation of the Cap domain from the open to the closed conformation. The represented ligands highlight the changes in the surface area at the catalytic pocket between the open and the closed conformations. The MG ligand shown in the apo-MpgP catalytic pocket was obtained from the refined coordinates of the [Mg^{2+}-MG-HPO_4^{2-}] structure. The bound ligands in the catalytic pocket of holo-MpgP are also represented. MG is drawn as sticks with atom colors yellow for carbon and red for oxygen; Mg^{2+} and the axial water molecule are represented as a small green and red spheres respectively, and are barely visible in the closed conformation of holo-MpgP. Atomic charges and radii were calculated for the protein residues only, with the APBS v0.4.0 PyMol plug-in (Baker, 2001). The electrostatic distribution was determined by applying the non-linear Poisson-Boltzmann method, considering T = 300 K, an ionic strength of 0.15 M, a solvent dielectric constant ε_s = 80 and a protein dielectric constant ε_p = 6. The molecular surface was calculated using a solvent probe radius of 1.4 Å. The range of electrostatic potentials shown spans from -5 (red) to +5 kT/e [1] (blue) units. Figure prepared with PyMOL (http://www.pymol.org).

[1] APBS writes out the electrostatic potential in dimensionless units of $k_B T/e$, where k_B is Boltzmann's constant ($1.3806504 \times 10^{-23}$ J K^{-1}), T is the temperature of the calculation in K and e is the charge of the electron ($1.60217646 \times 10^{-19}$ C).

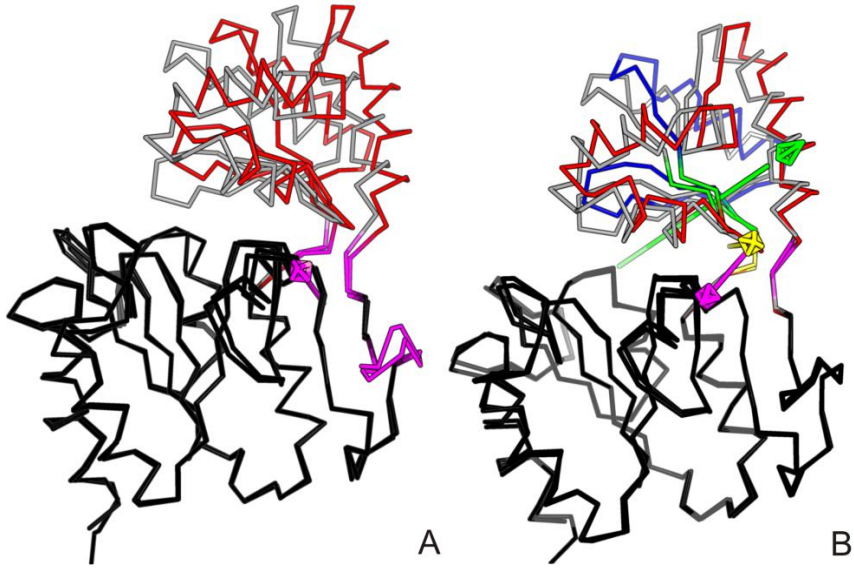


Figure 3.8 Hinge rotations of the C2B cap domain from the "open" to the "closed" conformations of MpgP The protein chains, represented by their C^{α} trace, were superposed at their HAD-like domains and illustrate the hinge rotation of the Cap domain from the "open" to the "closed" conformation. The HAD-like domains are colored black, and the C2B cap domain of holo-MpgP used as reference for the "closed"conformation is colored gray. (**A**) Superposition of "open" apo-MpgP (chain A) onto "closed" holo-MpgP (chain B). The entire C2B cap domain (red) rotates by about 11° about the hinge axis (magenta arrow). (**B**) Hinge rotations of "open" apo-MpgP (chain A) onto the "*quasi*-closed" conformation of holo-MpgP (chain A). Two flexible protein segments are distinguished at the C2B cap: the helical insertion motif (helices α4 and α5, residues Ala86−Glu139, red) and the Cof core region (residues Thr140−Ala172, blue). The colored arrows indicate the hinge rotation axes for the helical insertion motif (magenta, 7°; yellow, 9°) and for the Cof core region (green, 3°). The location of the bending and mechanical hinge residues is color-coded according to the color of their respective rotation axis. This figure was prepared with PyMOL (http://www.pymol.org).

3.5 DISCUSSION

Structural comparison of T. thermophilus MpgP with other members of the MPGP family

The different conformations found in the *T. thermophilus* MpgP crystal structures represented by their apo-MpgP chain A ("open") and holo-MpgP chains A ("*quasi*-closed") and B ("closed"), were compared by Secondary Structure Matching (SSM) (*44*), with their corresponding states in the orthologous MpgP crystal structures from *P. horikoshii* (PDB 2zos and 1wzc) and the MpgP-related YedP from *E. coli* (PDB 1xvi) (Fig. 3.9). Two distinct conformations ("open" for chain A and "closed" for chain B) were seen in the crystal structure of the *P. horikoshii* selenomethionine derivative (PDB 2zos), which corresponded to the apo-form (*8*). The "closed" conformation in the apo-form differed from that observed the in the native holo-MpgP (PDB 1wzc), in the "squiggle" and "flap" structural motifs as well as in Loop 3. These differences most likely arose from the co-crystallization of the native protein with magnesium and phosphate which resulted in a different crystal packing (*8*). Also, a low energy barrier between the alternative conformations may have facilitated the apo "closed" state in *P. horikoshii* MpgP.

The superposition between the "open" state of *T. thermophilus* MpgP with its orthologues gave overall mean r.m.s. C^{α} deviations of 2.20 Å for 1xvi and 1.96 Å for 2zos (chain B), whereas for the "closed" state the overall mean r.m.s. C^{α} deviations were 1.90 Å for 1wzc and 1.95 Å for 2zos (chain A). No structural homologue was found in the "*quasi*-closed" conformation. The superposition of the individual HAD-core and C2B cap domains gave slightly better results. For the "open" state, the superposition of the HAD-core domains of 1xvi and 2zos (chain B) with *T. thermophilus* MpgP gave mean r.m.s. C^{α} deviations of 1.95 Å and 1.66 Å respectively, while their C2B cap domains superposed with mean r.m.s. C^{α} deviations of 1.87 Å for and 1.53 Å respectively; for the "closed" state, mean r.m.s. C^{α} deviations of 1.30 Å and 1.50 Å were obtained respectively for the HAD-core domain superposition of

1wzc and 2zos (chain A) with *T. thermophilus* MpgP, whereas the superposition of the C2B cap domains gave mean r.m.s. C^{α} deviations of 1.85 Å for 1wzc and 1.35 Å for 2zos (chain A). These relatively high values reflect the fact that although the HAD-core and C2B cap domains sharing common folds in these structures, there are significant local differences arising from sequence insertions and deletions such as the relative orientation of α-helices and β-strands, as well as the position of loops such as the "squiggle" and "flap" flexible motifs of the HAD-core region.

In the "open" state of *P. horikoshii* apo-MpgP, Loop 3 is positioned in the plane of strands β8 and β9 while in *T. thermophilus* apo-MpgP and in YedP it follows the hinge motion of the Cap domain and is bent towards the catalytic pocket. This may result from the phosphate and sulfate ions bound to Loop 3 in MpgP and YedP respectively, as well as provide an explanation for the different hinge rotations observed for the *T. thermophilus* (11º) and *P. horikoshii* (19º) (*8*) enzymes. In the "closed" state, the hinge bending of the Cap domain moves Loop 3 further down into a conformation which is equivalent in both *P. horikoshii* and in *T. thermophilus* MpgP (*8*).

The catalytic pocket and its comparison with the other enzymes of the MPGP family

The catalytic pocket is located at the C2B and HAD-like interdomain region forming a binding cleft surrounded with ionizable and polar residues which are involved in substrate recognition and binding (Fig. 3.7 and Fig. 3.10). The four sequence signature motifs present in nearly all the aspartate-nucleophile HAD-SF members are placed at stereospecific regions of the HAD-core domain and are known to form a trigonal bipyramidal electrostatic mold that efficiently stabilizes negatively charged pentavalent transition state species (*2, 4*) (Fig. 3.5). Motif I is represented by Asp6-X-Asp8 in *T. thermophilus* MpgP, Asp8-X-Asp10 in *P. horikoshii* MpgP and Asp13-X-Asp15 in *E. coli* YedP, and is located at the C-terminal end of strand β1. Motif II consists of Thr39 (Ser40 in *P. horikoshii* MpgP and Ser47 in *E. coli* YedP) and

is placed at the C-terminal end of strand β2. Motif III is sited at the N-terminus of helix α7 and contains Lys177 (Lys180 in *P. horikoshii* MpgP and Lys188 in *E. coli* YedP) oriented towards Motifs I and II. Motif IV consists of the acidic signature Gly201-Asp202-Ser203-X-Asn205-Asp206 (Gly203-Asp204-Ser205-X-Asn207-Asp208 in *P. horikoshii* MpgP and Gly213-Asp214-Gly215-X-Asn217-Asp218 in *E. coli* YedP), that flanks the 3_{10} helical turn between strand β10 and helix α9.

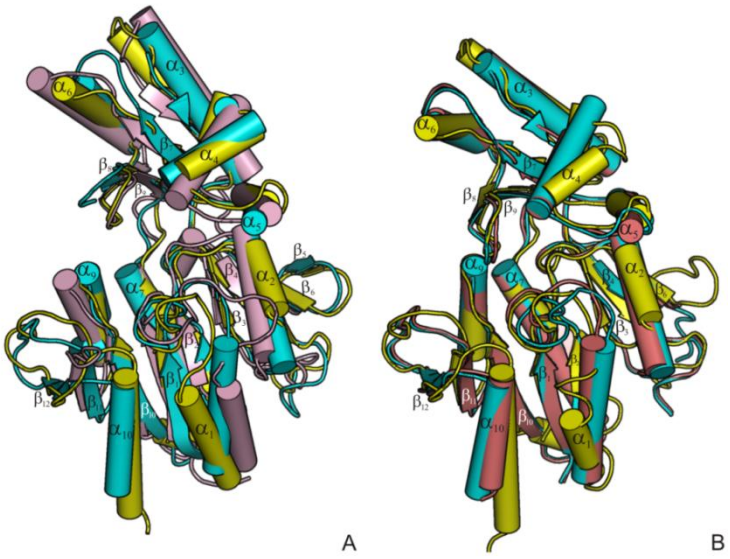


Figure 3.9 Structural comparison of *T. thermophilus* MpgP with its orthologous members of the MPGP family. **A)** Superposition of apo-MpgP (chain A, open conformation, in yellow) with *E. coli* YedP (PDB 1xvi, chain A, in light pink) and *P. horikoshii* apo-MpgP selenomethionine derivative (PDB 2zos, chain B, in cyan). **B)** Superposition of control [Mg^{2+}-MG-HPO_4^{2-}] structure (chain B, closed conformation, in yellow) with *P. horikoshii* native holo-MpgP (PDB 2wzc, chain B, in deep-salmon) and with its apo-MpgP selenomethionine derivative (PDB 2zos, chain A, in cyan). Secondary structure elements are numbered according to Fig. 3.5.

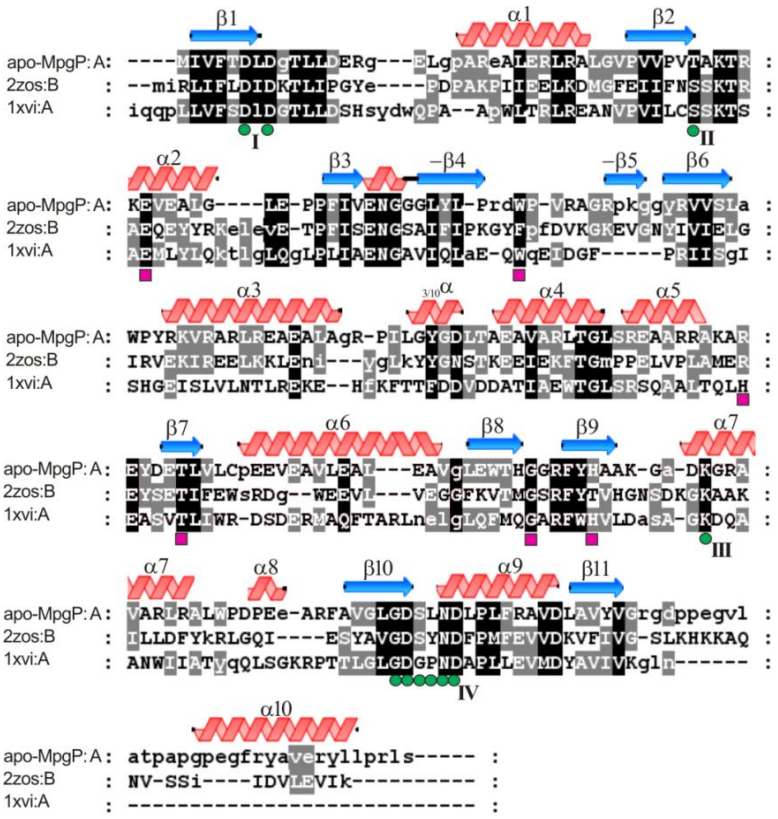


Figure 3.10 Structure-based sequence alignment of MpgP with its structural homologues. *T. thermophilus* HB27 apo-MpgP (chain A) was aligned with its closest homologues of the MPGP phosphatase family, in their correspondent apo conformations: *P. horikoshii* MpgP (PDB 2zos, chain B) and *E. coli* YedP (PDB 1xvi, chain A). Sequence alignment was done with ClustalW v.2.0.10, using the default parameters and edited with GENEDOC v.2.0 (http://iubio.bio.indiana.edu/) in agreement with a Secondary Structure Matching calculation in COOT (*34, 44*). Secondary structure elements correspond to the MpgP reference structure and are labelled according to Fig. 3.5. Helices are shown as pink coils and strands are represented by blue arrows. Conserved catalytic residues are marked according to their binding specificity: green circles indicate the positions of the conserved Motifs I to IV of the catalytic HAD-core region, while pink rectangles depict the residues involved in MG recognition and binding.

The recognition and binding of the MG group is done by the interfacing residues arranged at the top and back-side of the catalytic pocket (Fig. 3.11 A). A diagram of the MG molecule along with the atom numbering scheme used in this work is displayed in Fig. 3.11B. The unbiased sigmaA-weighted $|F_o|$ - $|F_c|$ electron density for the ligands and water molecules in the [Mg^{2+}-MG-HPO_4^{2-}] control structure is displayed in Fig. 3.11C and the corresponding ligand electron density for the other complexes discussed in this work is included in Fig. 3.12 and Fig. 3.13. At the α-face of the mannosyl-moiety a hydrogen-bond network is established between the HAD-like core residues Lys41 (Lys42 in *P. horikoshii* MpgP and Lys49 in *E. coli* YedP) and Glu45 (Glu46 in *P. horikoshii* MpgP and Glu53 in *E. coli* YedP) with the hydroxyl groups of carbons C3, C3(OH), and C4, C4(OH), of the sugar ring. At the β-face, atoms from residues in Loop 3 (Arg167 N in MpgP, Arg170 N in *P. horikoshii* MpgP and Arg177 N in *E. coli* YedP), and strand β9 (His170 $N^{\varepsilon 2}$ in MpgP, Thr173 O^{γ} in *P. horikoshii* MpgP and His180 $N^{\varepsilon 2}$ in *E. coli* YedP) stabilize the carboxylate group of the glyceryl-moiety. These residues line the region of the inner pocket surface which molds an electrophilic environment for mannosyl-sugar recognition and binding. In particular, the side-chain of Arg167 plays a key role in fixing the mannosyl-moiety by making a bidentate hydrogen bond that bridges one of its terminal NH_2 groups to the hydroxyl group C6(OH) of the mannosyl-moiety, and to the main-chain carbonyl oxygen of Thr122 (Thr127 in *P. horikoshii* MpgP and Thr130 in *E. coli* YedP). This second interaction makes an important contribution to the closing of Loop 3. The carbonyl oxygen of Gly165 (Gly168 in *P. horikoshii* MpgP and Gly175 in *E. coli* YedP) also accepts hydrogen bonds from the hydroxyl oxygen O13, hereafter referred to as O13(H), and a water molecule coordinating the cocatalytic metal center. The catalytic roles of Gly165 and Arg167 will be further discussed in the following sections. Tyr110 (Tyr115 in *P. horikoshii* MpgP and Phe118 in *E. coli* YedP) is located at the interface of the two C2B cap sub-domains. Tyr110 OH, together with Thr140 OH (Thr145 OH in *P. horikoshii* MpgP and Thr148 OH in *E. coli* YedP), stabilizes the hydroxyl group from the C2 epimer, C2(OH), and both are likely

to assist in the stereochemical discrimination of the mannosyl-group from other sugar-compounds. The C3(OH) hydroxyl of the mannosyl-moiety in also anchored by Arg135 $N^{\zeta 1}$ (Arg140 in *P. horikoshii* and His143 in *E. coli* YedP). The shorter side-chain of His143 in YedP strengthens the notion that it is a MpgP-related enzyme which is likely to use as substrate a modified sugar-compound at carbon C3. Interestingly, the ability of *T. thermophilus* and *P. horikoshii* MpgPs to use glucosyl-3-phosphoglycerate (GPG) as substrate, in addition to MPG (*54*), can be explained by the lack of steric hindrance towards the C2(OH) of the glucopyranosyl ring.

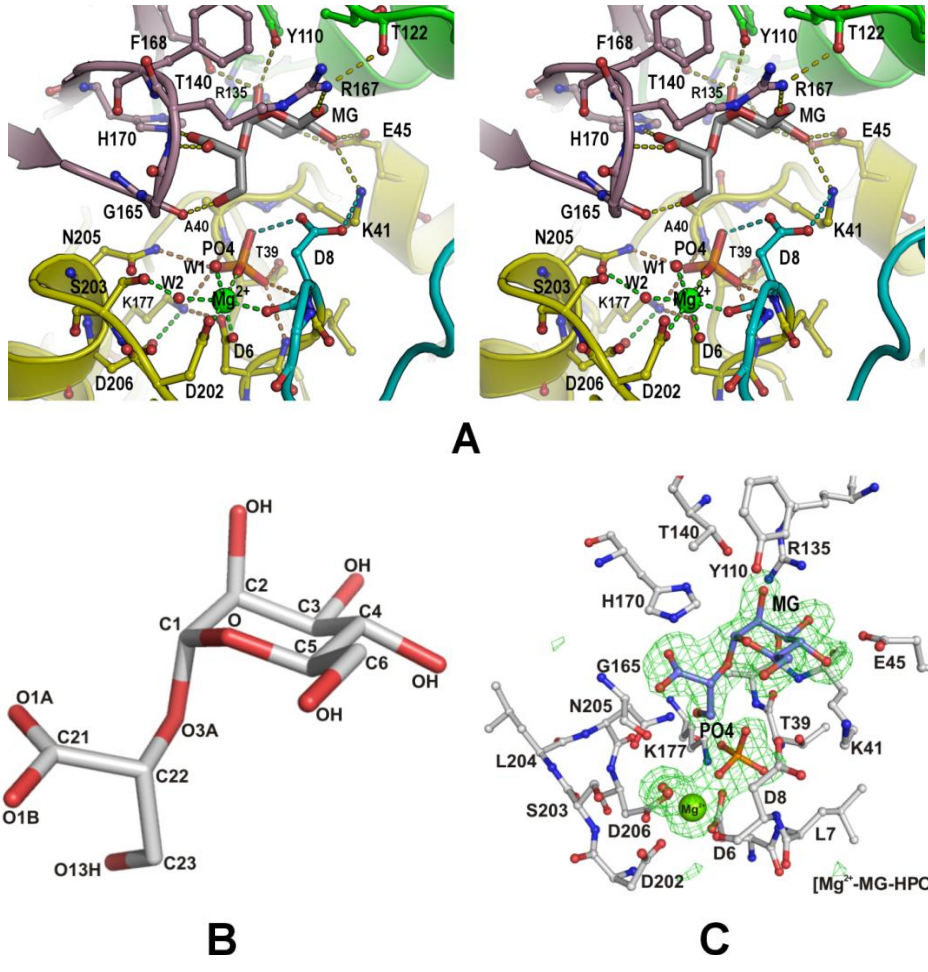


Figure 3.11 The MpgP catalytic pocket. A) Stereoview of the MpgP active site. The ternary complex bound in the "closed" conformation of the control [Mg^{2+}-MG-HPO_4^{2-}] structure is represented. Residues involved in the recognition and binding of either the MPG substrate or the MG and HPO_4^{2-} products, are represented in ball-and-stick, the HPO_4^{2-} and MG ligands are displayed as sticks and the water molecules and the Mg^{2+} ion are shown as spheres. Dashed lines indicate hydrogen bonds and are colour-coded according to their binding pocket regions: green for the cocatalytic metal site, orange for the phosphate binding site and yellow for the MG binding site. Atom colors are: green for magnesium, orange for phosphorus, gray for MG carbon atoms, blue for nitrogen and red for oxygen. The carbon atoms of MpgP are colored according the structural regions: yellow for the HAD-like region, cyan for the "squiggle" and "flap" motifs, green for the helical insertion α4 and α5 of the C2B cap module and light pink for the Cof core motif. **B)** Representation of the MG ligand. MG is oriented similarly as displayed in panel A. The carbon and oxygen atoms are labeled according to the nomenclature used in the text. **C)** View of the catalytic pocket in chain B of the control crystal structure [Mg^{2+}-MG-HPO_4^{2-}], showing the unbiased sigmaA-weighted $|F_o|$ - $|F_c|$ electron density map covering the ligands. The map was obtained with PHENIX by simulated annealing after omitting the ligands and the water molecules coordinated to the co-catalytic Mg^{2+} metal ion. The map is represented as a green mesh contoured at the 3 σ level. Protein residues and the MG ligand are shown in ball-and-stick representation and the Mg^{2+} metal ion and the water molecules are drawn as spheres. Atom colors are green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, white for protein carbon atoms and slate blue for MG carbon atoms. Figure prepared with PyMOL (http://www.pymol.org).

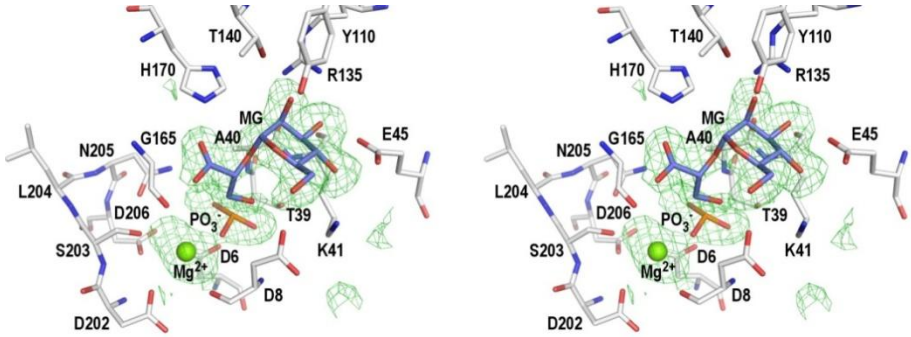
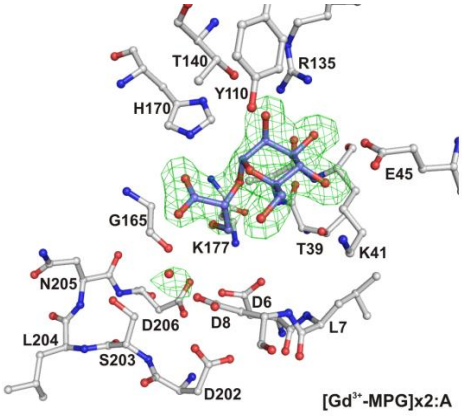
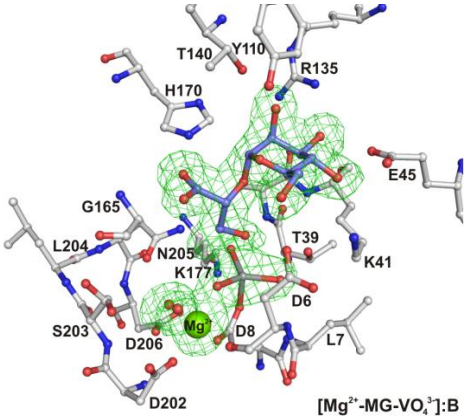
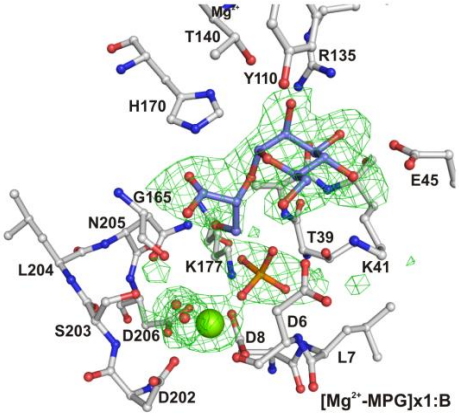
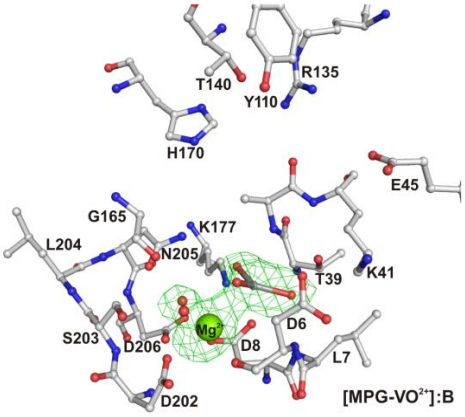


Figure 3.12 Stereoview of the [Gd^{3+}-MPG] crystal 1 structure showing the formation of the PO_3^- intermediate. The sigmaA-weighted $|F_o|$ - $|F_c|$ electron density map covering the ligands in the [Gd^{3+}-MPG] crystal 1 structure, a crystallographic snapshot of the bound ternary complex MpgP [Mg^{2+}-MG-PO_3^-], obtained with PHENIX by simulated annealing in the absence of the ligands, is represented as a green mesh and contoured at the 2 σ level. A trapped mixture of partially hydrolyzed MPG substrate shows the formation of a PO_3^- intermediate, as well as some disorder at the cocatalytic metal center. Atom colors are: green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, white for protein carbon atoms and slate blue for MG carbon atoms. Figure prepared with PyMOL (http://www.pymol.org).


T140
R135
Y110
H170
E45
G165
K177
T39
K41
N205
D6
D206
D8
L7
L204
S203
D202
[Gd3+-MPG]x2:A


T140
Y110
R135
H170
E45
G165
T39
K41
L204
N205
K177
D6
Mg2+
S203
D206
D8
L7
D202
[Mg2+-MG-VO4 3-]:B


Mg2+
T140
R135
Y110
H170
E45
G165
N205
T39
L204
K177
K41
S203
D206
D8
D6
L7
D202
[Mg2+-MPG]x1:B


T140
R135
Y110
H170
E45
G165
K177
L204
N205
T39
K41
Mg2+
S203
D206
D8
D6
L7
D202
[MPG-VO2+]:B

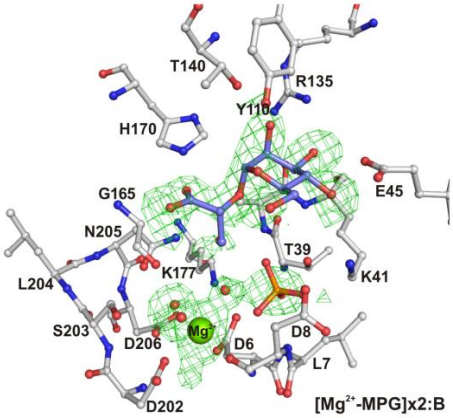

T140
R135
Y110
H170
E45
G165
N205
T39
K177
K41
L204
S203
Mg2+
D206
D6
D8
L7
D202
[Mg2+-MPG]x2:B

Figure 3.13 Unbiased sigmaA-weighted $|F_o|$ - $|F_c|$ electron density map covering the ligands in the catalytic pockets of structures [Gd^{3+}-MPG] crystal 2, [Mg^{2+}-MPG] crystal 1 and crystal 2, [MPG-VO^{2+}] and the control structure [Mg^{2+}-MG-VO_4^{3-}]. In each panel, the corresponding crystal structure and the represented MpgP chain are indicated at the bottom right. The $|F_o|$ - $|F_c|$ electron density maps were obtained with PHENIX by simulated annealing in the absence of the ligands and water molecules located inside the catalytic pocket and coordinated to the co-catalytic Mg^{2+} metal ion (when present). For each structure, the maps are represented as a green mesh and contoured at the 3 σ level for the structures [Gd^{3+}-MPG] crystal 2, [Mg^{2+}-MG-VO_4^{3-}] and [MPG-VO^{2+}], and at the 2 σ level for the structures [Mg^{2+}-MPG] crystal 1 and crystal 2. Protein residues and polyatomic ligands are drawn in ball-and-stick representation. Atom colors are green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, white for protein carbon atoms and slate blue for MG carbon atoms. Figure prepared with PyMOL (http://www.pymol.org).

Coordination of the cocatalytic metal-ion binding site

In all crystal structures except [Gd^{3+}], a coordinated Mg^{2+} metal ion was refined at the cocatalytic metal binding site. This happened even with the [Gd^{3+}-MPG] crystal 1 and 2 structures, soaked with $GdCl_3$ instead of $MgCl_2$, which was prepared with the aim of trapping a reaction coordinate mimic (Table 3.2). However, no anomalous signal was detected for Gd^{3+} and the electron density peak present at the cocatalytic metal binding site was also interpreted as a Mg^{2+} metal ion, probably present in trace amounts in the soaking solution and with high affinity for the metal site. This may also explain the presence of an Mg^{2+} ion in the [MPG-VO^{2+}] crystal structure. The binding of the cocatalytic Mg^{2+} ion should follow a structural rearrangement of the "squiggle" and "flap" motifs, as suggested by their different conformations in chains A and B of the apo-MpgP, and in chain A of the metal-apo form in the [Gd^{3+}-MPG] crystal 2 structure (Fig. 3.14 and Fig. 3.15). Thus enzyme

activation implies the binding of the cocatalytic Mg^{2+} ion for the subsequent phosphoryl-transfer process.

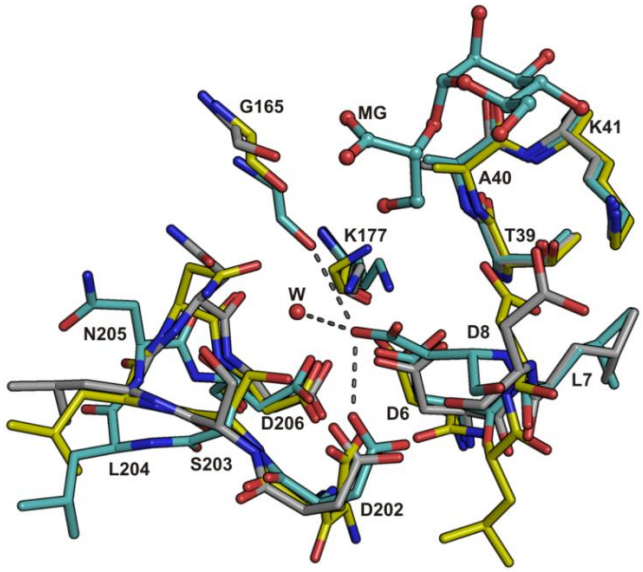


Figure 3.14 Alternative conformations of the catalytic HAD-core in the apo-state. The structures are superposed at their HAD-like domains and illustrate the structural variability of the apo-state in the catalytic HAD-core region. The inactive conformation is represented by the "open" apo-MpgP (chain B, C^{α} atoms in yellow), and by the apo-HAD catalytic core of the "*quasi*-closed" [Gd^{3+}-MPG] crystal 2 structure (chain A, C^{α} atoms in cyan), which illustrates the likelihood of Asp8 being protonated in the apo-state, by showing the interactions (grey dashed lines) between Asp8 $O^{\delta 2}$ with three neighbouring oxygen atoms: O from water W, Gly165 O and Asp202 $O^{\delta 1}$. Nevertheless, the conserved salt bridge between Asp8 and Lys41, present in the active state of the enzyme, is also formed in the apo-MpgP (chain A, C^{α} atoms in grey), due to the high mobility of the "squiggle" and "flap" motifs in the crystal structure. The MG leaving group indicates the MG binding site and Gly165 illustrates the closure motion of the C2B cap domain. Figure prepared with PyMOL (http://www.pymol.org).

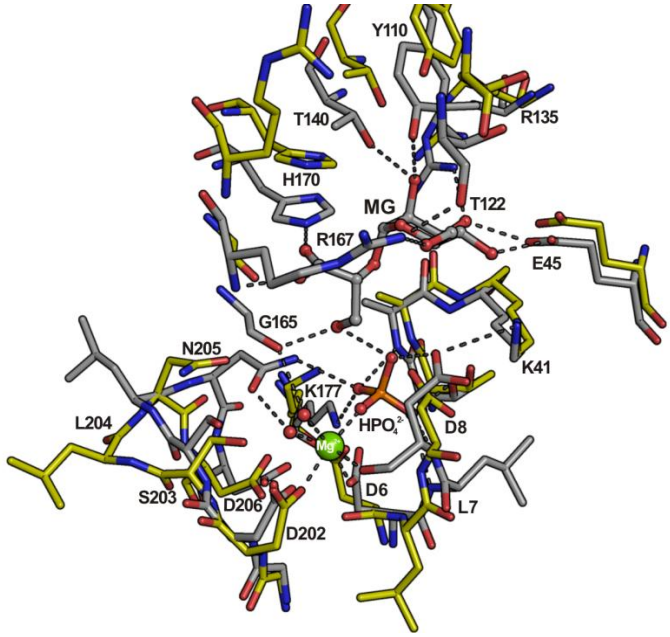


Figure 3.15 Enzyme activation and closure via hinge bending. Overlay of the active sites of apo-MpgP chain B and [Mg^{2+}-MG-HPO_4^{2-}] control structure chain B, which correspond to the open and inactive state and the closed and active state of the enzyme, respectively. The structures are superposed at their HAD-like domains and illustrate the changes occurring in the HAD core region upon the binding of the cocatalytic metal ion Mg^{2+} and the phosphate moiety of MPG (represented by the ligand HPO_4^{2-}). Residues involved in the recognition and binding of either the MPG substrate or the MG and HPO_4^{2-} products are drawn as sticks, whereas the HPO_4^{2-} and MG ligands are shown in ball-and-stick representation. Water molecules and the Mg^{2+} ion are displayed as spheres. Hydrogen bonds established between ligands and holo-MpgP, and relevant for hinge bending, are represented as dashed lines. Atom colors are as follows: green for magnesium, orange for phosphorus, gray for MG carbon atoms, blue for nitrogen, and red for oxygen. The carbon atoms of MpgP are colored yellow for apo-MpgP and gray for holo-MpgP. This figure was prepared with PyMOL (http://www.pymol.org).

Binding and stabilization of the phosphoryl species

The binding modes of the orthophosphate observed in the chain B of the control crystal structure [Mg^{2+}-MG-HPO_4^{2-}] and in chain B of [Mg^{2+}-MGP] crystal 1 structure are similar. Three oxygen atoms are stabilized by the amide hydrogen atoms from the Motif I peptide backbone (Leu7 N, Asp8 N) and the neighboring residue Ala40 (Ala40 N) in Motif II, the cocatalytic Mg^{2+} metal-ion, Lys177 N^{ζ} in Motif III and Asn205 $N^{\delta 2}$ from Motif IV (Fig. 3.11A). The fourth oxygen of HPO_4^{2-} accepts hydrogen bonds from the hydrolyzed ester oxygen O13(H) of the MG leaving group and Asp8 $O^{\delta 2}$. These residues form an electrostatic mold with a trigonal bipyramidal-like geometry that assists in the stabilization of a developing TS species (*4*).This geometry is illustrated by the aspartyl-vanadate adduct of control crystal structure [Mg^{2+}-MG-VO_4^{3-}] (Fig. 3.16B, right panel). The vanadium atom is equatorially coordinated by three oxygen atoms (restrained bond lengths of 1.9 Å) and axially by Asp6 $O^{\delta 1}$ and MG O13 with V—O distances of 2.0 Å and 2.7 Å respectively, in agreement with previous results (*4*). In this structure, the phosphoryl-analogue orthovanadate (VO_4^{3-}) underwent an equatorial polarization with expansion of the axial V—O bond length, prior to its nucleophilic attack by Asp6 $O^{\delta 2}$ with subsequent formation of an aspartyl-vanadate adduct. However, this may not truly reflect the activation chemistry for phosphoryl-compounds (*55*). Indeed, and contrary to the stronger additive character required for the formation of a phosphoryl-adduct intermediate as expected by the associative mechanism of phosphoryl-transfer (*55*), the electron density of the transferring "PO_3^--species" in the active site of the [Gd^{3+}-MPG] crystal 1 structure suggests a mixture of non-hydrolyzed R-PO_4^{2-} and hydrolyzed "PO_3^--species" (hereafter designated as "PO_3^--like intermediate" (Fig. 3.12), reinforcing the notion of a concerted dissociative-like mechanism in this first step, with weaker bond order formation from the attacking nucleophile Asp6. *In vivo*, a short-lived metaphosphate (PO_3^-) intermediate is very likely stabilized by the trigonal bipyramidal electrostatic mold described above as suggested by the [Gd^{3+}-MPG] crystal 1 structure (Fig. 3.16B, middle panel). The phosphorus center should become

positively charged and stabilized by the hydrolyzed ester-oxygen of the MG leaving group, either as an alkoxide anion C23(O13^{-}) or as a hydroxyl group O13(H), and by Asp6 O$^{\delta 1}$. These two oxygen atoms are axially located with relation to the plane defined by the three *pseudo*-equatorial oxygen atoms of the "PO_3^{-}-like intermediate". These two atom groups are hereafter referred to as "non-bridging" and "bridging" oxygen atoms, respectively. In contrast with the salt bridge between Asp10 and Arg170 in the holo-MpgP structure of *P. horikoshii* (*8*), a salt bridge between corresponding residues Asp8 and Arg167 is not observed in the *T. thermophilus* MpgP structures. This salt bridge was suggested to prevent the reverse phosphoryl-transfer reaction (*8*), yet such role is doubtful due to the absence of a true Michaelis complex bound at the active site. Instead, in the *T. thermophilus* MpgP structures, Asp8 interacts with Lys41 and assumes a conformation similar to that observed for several members of the IIB and III HAD-subfamilies (*4*) (Fig. 3.11A). In the [Gd^{3+}-MPG] crystal 1 structure, Asp8 O$^{\delta 2}$ is oriented in-line to O13(H) of the MG leaving group, and makes a bifurcated hydrogen bond with O13(H) (2.5 Å) and with the amine group of Lys41 (2.8 Å) (Figure 3.16B, middle panel). These interactions highlight the role of Asp8 as a general acid catalyst. On the other hand, in the [Mg^{2+}-MPG] crystal 1 structure, when the "bridging" oxygen atoms move away from their axial positions, a salt bridge between Asp8 and Lys41 is likely to be formed (Figure 3.16D, middle panel). In the [Mg^{2+}-MPG] crystal 2 structure the HPO_4^{2-} ion is found closer to the pocket exit, suggesting a HPO_4^{2-} release concerted with the collapse of the trigonal bipyramidal electrostatic mold, primarily caused by the absence of the axial stabilization from Asp8 O$^{\delta 2}$ (Figure 3.16E, middle panel).

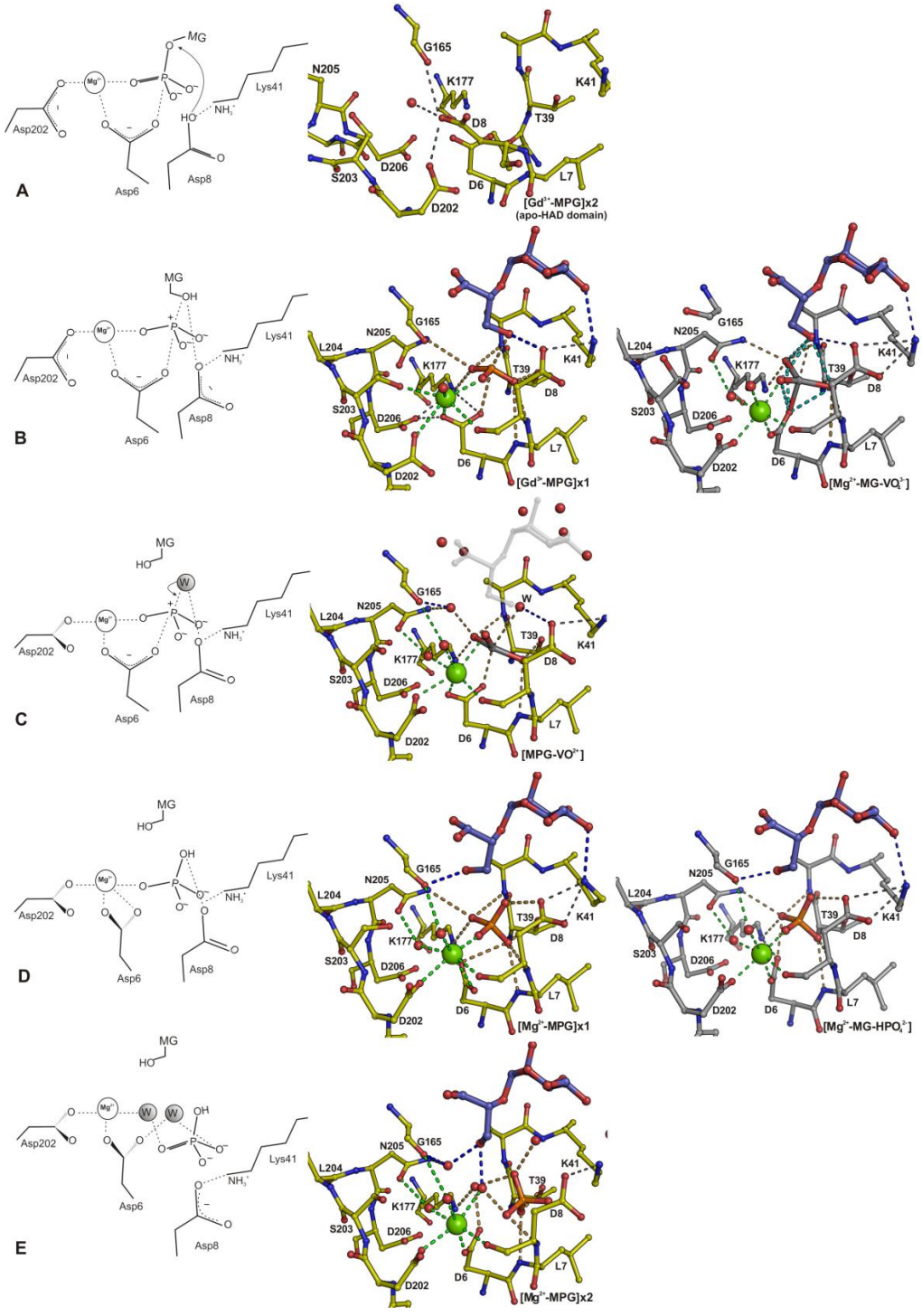

A
MG
Mg2+
Asp202
Asp6
Asp8
Lys41
G165
N205
K177
K41
D8
T39
S203
D206
D6
L7
D202
[Gd3+-MPG]x2
(apo-HAD domain)
B
MG
OH
L204
[Gd3+-MPG]x1
[Mg2+-MG-VO4 2-]
C
W
[MPG-VO2+]
D
[Mg2+-MPG]x1
[Mg2+-MG-HPO4 2-]
E
[Mg2+-MPG]x2

Figure 3.16 The catalytic events in the proposed phosphoryl-transfer mechanism for MpgP. Except for the snapshot in **A)**, for each event a diagram is shown to the left, with its corresponding crystallographic snapshot (**B, D, E**) or mimicking "PO_3^--intermediate" structure (**E**) in the middle, and the respective control crystal structure (**B, D**) on the right. In each diagram, only key residues involved in the synchronous migration of the cocatalytic Mg^{2+} ion with the transferring "PO_3^--species", and in the assisted acid catalysis are depicted. For clarity, minimal electrostatic interactions between the catalytic residues with the migrating species are shown as dashed lines, while arrows indicate the assisted proton-transfer mechanism and nucleophilic attack by the catalytic water molecule. In the middle and left panels, the electrostatic interactions between the catalytic HAD-core residues and the migrating Mg^{2+} and "PO_3^--species" are displayed as dashed lines and are color coded according to their respective binding sites; green for the cocatalytic Mg^{2+} binding site and bright orange for the phosphoryl binding site, while those extending towards the MG binding site are shown in blue. Electrostatic interactions between protein residues, as well as hydrogen bonds between protein residues and water molecules are shown as grey dashed lines. **A)** active enzyme, following the structural rearrangement of the HAD-core residues in concert with the binding of the cocatalytic Mg^{2+} ion and the phosphoryl group of the MPG substrate. The Asp8-Lys41 salt bridge positions Asp8 in-line to the ester-oxygen for the assisted acid catalysis. In the middle panel, the [Gd^{3+}-MPG] crystal 2 structure illustrates one of the possible conformations acquired by the HAD-core region in its apo-form, and the ability for Asp8 to become protonated. **B)** After cleavage of the phosphoester bond, the short lived "PO_3^-- intermediate" is stabilized by the trigonal-bipyramidal electrostatic mold, as illustrated by the [Gd^{3+}-MPG] crystal 1 structure in the central panel. In the right panel, the aspartyl-vanadate adduct of the [Mg^{2+}-MG-VO_4^{3-}] crystal structure is stabilized by the trigonal bipyramidal electrostactic mold in an ideal geometry. **C)** nucleophilic attack at the PO_3^- center by a water molecule, represented by the mimicking VO_3^- complex and its axial water

molecule, W. **B)** to **D)** side-chain rotation of the residues Asp6 and Asp202, in concert with the migration of the Mg^{2+} and "PO_3^--species" in **B)** or mimicking [Mg^{2+}-VO_3^-] complex in **C)**, towards their final positions as in **D)**. **E)** HPO_4^{2-} release followed by the collapse of the trigonal bipyramidal electrostatic mold. Atom colors are: green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, yellow for protein carbon atoms and slate blue for MG carbon atoms. Figure prepared with PyMOL (http://www.pymol.org).

Structural changes in the MG binding pocket during MPG binding and catalysis

The anchoring of the phosphoryl-group of the MPG substrate to the cocatalytic Mg^{2+} ion favors the hinge bending closure motion and the shaping of the catalytic pocket, fitted towards the recognition and binding of the sugar moiety (Fig. 3.15). This scaffold is also found for chain B in the crystal structure of the metavanadate complex [MPG-VO^{2+}] where neither bound MPG nor MG were observed. Instead, water molecules are hydrogen-bonded to the same residues that interact with the hydroxyl groups of the sugar substrate (Fig. 3.16C, middle panel). Locally, residues in Loop 3 are also key mediators in the reaction cycle while binding to the sugar substrate. In particular, Arg167 should play a central role in the global bending of the C2B cap module by stabilizing the mannosyl-moiety *via* the C6(OH) group as previously described.

In the [Mg^{2+}-MPG] crystal 1 structure, the hydroxymethyl group C23O13(H) rotates by 108º about its σ-bond to the preceding carbon atom, forming hydrogen bonds with both Gly165 O (Fig. 3.16D, middle panel). On the other hand, in the [Mg^{2+}-MG-VO_4^{3-}] crystal structure the O13(H) group of MG is still hydrogen-bonded with Asp8 $O^{\delta 2}$ and axially coordinates the vanadium atom (Fig. 3.16B, right panel). Therefore, the enzyme-inhibited state obtained by the formation of the aspartyl-vanadate adduct in the [Mg^{2+}-MG-VO_4^{3-}] crystal structure mirrors the acid-assisted step of phosphoryl-transfer catalysis. Gly165 O should play a primary role in the second step, by stabilizing the conformational change of the C23O13(H) group observed in the [Mg^{2+}-MPG]

crystal 1 structure, and thus allowing the nucleophilic attack of the PO_3^- center by an "activated" water molecule. Indeed, in the [Mg^{2+}-MG-VO_4^{3-}] crystal structure, the Gly165 carbonyl is flipped, pointing in the opposite direction from that found in all the other *T. thermophilus* MpgP crystal structures herein presented (Fig. 3.16B, right panel), thus hinting for its catalytic relevance.

The phosphoryl-transfer mechanism in MpgP

During the crystal soaking procedures, the crystallized MpgP displayed a catalytic rate slow enough to allow the trapping of hydrolyzed species at several stages of the phosphoryl-transfer reaction. These were rationalized as "crystallographic snapshots", reflecting a series of molecular and structural events that take place during the catalytic cycle. Additionally, the soaks done with the hydrolyzed products (MG and HPO_4^{2-}), and also in the presence of Vanadium (V) oxoanion compounds (VO_3^- and VO_4^{3-}), were used as control crystal structures, allowing a detailed analysis of the binding residues and of the extended electrostatic network that is engaged and drives protein dynamics into its catalytically competent structural states along the reaction coordinate. These structural results paralleled the "crystallographic snapshots" of MPG hydrolysis, thus supporting the insights into the derived catalytic mechanism herein presented.

The catalytic cycle of MPG hydrolysis is schematically represented in Fig. 3.16 and the key events of the reaction coordinate are interpreted on the basis of the structural data. In the [Gd^{3+}-MPG] crystal 1 structure the trapped PO_3^--like species exhibits an apparently asymmetric axial coordination, *i.e.*, 2.7 Å to the hydrolyzed ester-oxygen O13 and 2.9 Å to Asp6 $O^{\delta 1}$ (Fig. 3.16B, middle panel). This difference is of the same magnitude as the overall mean coordinate error estimate for this crystal structure (0.27 Å) and therefore its significance may be questioned. However, it is likely that the residues in the active site are more rigid and thus have lower coordinate errors than those on the surface of the enzyme. This asymmetry suggests that the phosphoryl-transfer reaction in MpgP should undergo a concerted dissociative-like

mechanism (A_ND_N), but with a weaker S_N2 character for the stabilization of the resonance species of the PO_3^- intermediate. In this view, the reactivity of the MG leaving group should overcome that of the attacking nucleophile Asp6, with a transition state where the bond to the leaving group is more extensively broken than the bond to the nucleophile is formed (*11, 16*). Higher resolution diffraction data would be needed to verify this hypothesis. At this stage, the MG leaving group remains fixed in its binding site, whereas the PO_3^--like species is stabilized by the trigonal bipyramidal electrostatic mold formed by Motifs I-IV. This is in agreement with previous studies done with the predicted hexose phosphate phosphatase (HPP) from *Bacteroides thetaiotaomicron* (*4*)*,* in which the role of this mold was assessed in connection with the stabilization of the trigonal bipyramidal TS species and the catalytic turnover. A di-anionic monoester-phosphate should be the reactive form of the MPG substrate, as suggested by the pH dependency of the MpgP catalytic activity, with its highest turn-over rate at about pH 6.0 (*1, 16*). This fact highlights the need for an external acid catalyst, required to enhance the rate of phosphoryl-transfer (*4, 56*), contrary to the intramolecular proton shift postulated for mono-anions of phosphate monoesters (*16, 57*).The [Gd^{3+}-MPG] crystal 1 structure represents the acid-assisted catalysis in the first step of the phosphoryl-transfer mechanism (Fig. 3.16B, middle panel). Asp8 is predicted to act as the proton donor, supported by the hydrogen bond between the hydrolyzed ester-oxygen O13 and Asp8 $O^{\delta2}$ (2.5 Å), together with the trapped "PO_3^--like intermediate" and implying the coupled protonation of the hydroxymethyl anion $C(23)O(13)^-$ with the P—O bond fission. Asp8 can be transiently protonated prior to the binding of the cocatalytic metal ion (Fig. 3.15A). This is illustrated by the [Gd^{3+}-MPG] crystal 2 structure, where in chain A Asp8 is hydrogen-bonded to the carbonyl oxygen of Gly165 (Fig. 3.14). Indeed, this structure is a "snapshot" of the final stages of the catalytic cycle: in chain B, the HPO_4^{2-} product has been replaced by two water molecules, followed by a destabilization of the cocatalytic metal center, whereas in the "*quasi*-closed" conformation of chain A, the labile cocatalytic metal ion is released, allowing Asp8 to receive a proton

from a water molecule of the solvent channel that fills the metal binding site. On the other hand, the energetic barrier for proton donation is expected to be significantly lowered, both by the relatively high pK_a predicted for the MG leaving group (*16, 56*) as well as the electrostatic effect between Asp8 and Lys41. In the latter, the observed salt bridge which has been proposed for several members of the IIB and III HAD-subfamilies (*58*) will have the effect of lowering the pK_a of Asp8 and raising the pK_a of Lys41, as would be generally the case for a coupled acid-base pair (*59*), thus destabilizing the protonated state of Asp8 and favoring its proton donation to the ester-oxygen of MPG, while this is being stabilized for the "in-line" assisted acid catalysis (Fig. 3.16B).

The interaction between Asp6 $O^{\delta1}$ and the positive charge at the phosphorus atom in the PO_3^-- intermediate (Fig. 3.16B), facilitates P—O bond fission and contributes for the synchronous migration of the anchored "PO_3^--species" with the loosely bound cocatalytic Mg^{2+} ion, when comparing their positions with those in the [Mg^{2+}-MPG] crystal 1 structure, which represents the final reaction products (Fig. 3.16B and 3.16D). More significantly, this relatively small motion is accompanied by large changes in the orientation of the carboxyl groups of Asp6, and Asp202: both rotate by about 90º with respect to their orientation in the [Gd^{3+}-MPG] crystal 1 structure. In particular, the rotation of Asp6 seems to be directly related with the electrostatic interaction between Asp6 $O^{\delta1,2}$ and the phosphorus atom in HPO_4^{2-}, whereas the rotation of Asp202 appears to be connected with the coordination of Mg^{2+} by Asp202 $O^{\delta1}$ (Fig. 3.16C). Similar orientations were observed for the carboxylate group of Asp6 in the control crystal structures [Mg^{2+}-MG-VO_4^{3-}] and [Mg^{2+}-MG-HPO_4^{2-}], which replicate those of the crystal 1 structures of [Gd^{3+}-MPG] and [Mg^{2+}-MPG] crystal 1, respectively. In the control crystal structure [Mg^{2+}-MG-VO_4^{3-}], Asp6 $O^{\delta1}$ is bound axially to the Vanadium center (2.0 Å) (Fig. 3.16B, right panel), whereas in the control crystal structure [Mg^{2+}-MG-HPO_4^{2-}] Asp6 $O^{\delta2}$ interacts axially with a non-bridging oxygen atom of HPO_4^{2-} (2.8 Å), and has probably rotated due to electrostatic repulsion prior to the product release (Fig. 3.16D, right panel). Thus, the structural template provided by the active catalytic HAD-

core region, has the ability to hold in proper position the two migrating electrophilic centers, while balancing the hydrogen bonding network and electrostatic potential in their vicinity as needed to overcome the energetic barriers of P—O bond fission and P—O bond making. Indeed, the migration of the cocatalytic Mg^{2+} ion with the concomitant stereo-inversion of the PO_3^- intermediate, was previously proposed for crystals of fructose-1,6-bisphosphatase (FBPase), grown under an equilibrium mixture of substrates and products (*14*). Also, crystallographic snapshots of *Lactococcus lactis* β-phosphoglucomutase (β-PGM) revealed a migration of the cocatalytic Mg^{2+} center by 0.7 Å synchronously with the enzyme activation by phosphorylation, and in concert with a rotation of the carboxylate groups from motifs I and IV (*60*).

The second step of the phosphoryl-transfer reaction comprises the nucleophilic water attack to the PO_3^- intermediate, leading to the final products represented by the ligands in the end-reaction [Mg^{2+}-MPG] crystal 1 structure (Fig. 3.16C and 3.16D). The rotation of the hydroxymethyl group C23O13(H) away from Asp8 to form a hydrogen bond with Gly165 O leads to the exposure of the positive charge on the PO_3^- intermediate, which may then be readily attacked by a water molecule, followed by a proton release. In contrast with the associative phosphoryl-transfer mechanism of the C2B phosphatase HPP (*4*) where a role as general acid/base catalyst was proposed for the second Asp residue (Asp10), a similar role for Asp8 is questionable in MpgP. Its expected reduction in pK_a, which results from the salt bridge established with Lys41, favors a deprotonated state throughout the second step, and conflicts with Asp8 playing a role as a general base. Instead, Asp8 should mainly play a structural role during this stage by maintaining the trigonal bipyramidal electrostatic mold that holds the PO_3^- intermediate. Structural evidence for the preservation of the Asp8-Lys41 salt bridge during the second step of the reaction, is given by the PO_3^--mimicking intermediate [MPG-VO^{2+}] crystal structure, where an axial water molecule bridging Asp8 to the VO_3^- center mirrors either the hydrolyzed ester-oxygen O13(H) of MG or the attacking

water molecule (Fig. 3.16C), and by the apical interaction of Asp8 with the hydrolyzed HPO_4^{2-} product prior to its release (Fig. 3.16D). This may explain the lower catalytic rate measured for the HPP (D10A) mutant in the presence of a low energy leaving-group (*4*).

Table 3.1 Solutions used in the preparation of the [Gd^{3+}] crystals by soaking

Solution	[CarAc] *m*M	[$GdCl_3.xH_2O$] *m*M
1	95	5
2	50	5
3	25	5
4	10	5
5	2	5
6	2	10

[CarAc], total carboxylic acid concentration.

Table 3.2. Summary of crystallization and crystal soaking procedures for all datasets with their respective bound ligands [a].

Dataset	Resolution (Å)	Co-crystallization (ions)	Solution 5	Soak Time (hours)	Solution 6	Soak Time (hours)	Reaction products	Bound ligands
[Gd^{3+}]	2.50	Gd^{3+}	5 *m*M $GdCl_3$	N/A	10 *m*M $GdCl_3$	14	-	$GdCl_3.xH_2O$ [b]
apo-MpgP	1.80	Mg^{2+}, PO_4^{3-}	N/A	N/A	N/A	N/A	-	-
[Mg^{2+}-MPG] crystal 1	1.90	Mg^{2+}, PO_4^{3-}	5.8 *m*M MPG, 1 *m*M $MgCl_2$	5	5.8 *m*M MPG, 4 *m*M $MgCl_2$	1/4	MG, HPO_4^{2-}	Mg^{2+}, MG, HPO_4^{2-}
[Mg^{2+}-MPG] crystal 2	1.80	Mg^{2+}, PO_4^{3-}	5.8 *m*M MPG	6	10 *m*M MPG, 7 *m*M $MgCl_2$	1/3	MG, HPO_4^{2-}	Mg^{2+}, MG, HPO_4^{2-}
[Gd^{3+}-MPG] crystal 1	1.90	Gd^{3+}, PO_4^{3-}	5 *m*M MPG	1/3	6.7 *m*M MPG, 7.5 *m*M $GdCl_3$	1/2	-	Mg^{2+}, MG, PO_3^-
[Gd^{3+}-MPG] crystal 2	1.92	Gd^{3+}, PO_4^{2-}	5 *m*M MPG	1/3	6.7 *m*M MPG, 7.5 *m*M $GdCl_3$	20	MG	Mg^{2+}, MG
[Mg^{2+}-VO^{2+}]	2.10	Mg^{2+}, PO_4^{3-}	5 *m*M $VOSO_4$	22	10 *m*M MPG, 7 *m*M $VOSO_4$	1	$[VO_3]^-$ [c]	Mg^{2+}, VO_3^-(V)
[Mg^{2+}-MG-HPO_4^{2-}] (control)	1.74	Mg^{2+}, PO_4^{3-}	N/A	N/A	10 *m*M MG, 10 *m*M $NaKPO_4$, 7 *m*M $MgCl_2$	24	-	Mg^{2+}, MG, HPO_4^{2-}
[Mg^{2+}-MG-VO_4^{3-}] (control)	1.81	Gd^{3+}, PO_4^{3-}	N/A	N/A	10 mM MG, 10 mM Na_3VO_4, 7 mM $MgCl_2$	14	[VO_3(Asp 6 $O^{\delta 2}$)]$^{3-}$ [d]	Mg^{2+}, MG, [VO_3(Asp6 $O^{\delta 2}$)]$^{3-}$

[a] Only the substrates and/or substrates analogues added to the soaking solutions are described. The soaks comprised either one (Solution 6) or two steps (Solutions 5 and 6) depending on the dataset; [b] A gadolinium trinuclear cluster was refined in the crystal structure; [c] Oxidation of the vanadyl (IV) oxocation in solution to metavanadate (V) oxoanion: $VO^{2+} + O_2 \rightarrow VO_3^-$; [d] Nucleophilic attack of vanadium (V) oxoanion by Asp6 $O^{\delta 2}$ oxygen with subsequent formation of an aspartyl-vanadate adduct [VO_3(Asp6 $O^{\delta 2}$)]$^{3-}$. MG, α-D-mannosylglycerate; MPG, α-D-mannosyl-3-phosphoglycerate.

Table 3.3 Crystallographic data collection, processing and phase refinement statistics for the [Gd^{3+}] dataset

Data collection and processing statistics		*Phase refinement statistics*	
Wavelength (Å)	1.5418	Phasing power, anomalous	1.553
Space Group	$P2_1$	Anomalous R_{cullis}	0.692
Unit Cell Parameters (Å)	a = 39.65, b = 70.77, c = 92.68, β = 95.05°	SHARP FOM, acentric	0.467
Resolution range (Å)	39.49 - 2.50 (2.59 - 2.50)	SHARP FOM, centric	0.133
Scan type	ω-scan	SHARP FOM, overall	0.450
Total angular range (º) [a]	518.6	*Density modification statistics*	
Total number of frames [a]	2593	Overall $\|E^2\|$ correlation [e]	0.732
Exposure time per frame (seconds)	90	FOM after final DM run [e]	0.825
Nr. observations	195118 (12464)	FOM after NCS averaging [f]	0.786
Unique reflections	17882 (1833)		
<I/σ (I)>	17.2 (4.7)		
R-merge (%) [b]	10.3 (36.5)		
R-pim (%) [c]	3.2 (13.5)		
Completeness (%)	99.7 (98.2)		
Multiplicity	10.9 (8.1)		
Wilson B ($Å^2$)	38.0		
Nr. molecules in a.u.	2		
V_m ($Å^3$/Da) [d]	2.31		
Estimated solvent content (%) [d]	46.7		

Values in parentheses refer to the highest resolution shell. [a] In 8 different crystal settings; [b] R-merge = merging R-factor, $(\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|) / (\sum_{hkl} \sum_i I(hkl)) \times 100\%$; [c] R-pim = precision independent R-factor, $\sum_{hkl} [1/(N_{hkl} -1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl) \times 100\%$, where I is the observed intensity, $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections, and N_{hkl} is their multiplicity (*61*); [d] According to (*24*); [e] from SHARP optimizing density modification procedure (SOLOMON followed by final DM run) (*29*); [f] from second DM run (*30*).

Table 3.4 Data collection and processing statistics.

Data Sets	apo-MpgP	[Mg^{2+}-MPG] crystal 1	[Mg^{2+}-MPG] crystal 2	[Gd^{3+}-MPG] crystal 1	[Gd^{3+}-MPG] crystal 2	[MPG-VO^{2+}]	[Mg^{2+}-MG-HPO_4^{2-}] (Control)	[Mg^{2+}-MG-VO_4^{3-}] (Control)
Beamline	ID14-4	ID14-4	ID14-4	ID14-4	ID14-4	ID23-1	ID14-4	I03
Wavelength (Å)	0.9535	1.2520	0.9535	0.9535	0.9535	0.9792	1.2520	0.97625
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell parameters (Å,º)	a = 39.52 b = 70.68 c = 95.42 β = 92.95	a = 39.93 b = 70.53 c = 92.63 β= 95.70	a = 39.37 b = 88.05 c = 146.46	a = 39.79 b = 70.80 c = 97.77 β= 95.54	a = 39.87 b = 70.24 c = 91.78 β = 95.68	a = 39.97 b = 71.36 c = 92.97 β = 95.46	a = 39.90 b = 70.74 c = 92.30 β= 95.37	a = 39.95 b = 71.26 c = 91.22 β= 95.52
Resolution (Å)	50.00 - 1.90 (2.01 - 1.90)	50.00 - 1.90 (2.00 - 1.90)	60.00 - 1.80 (1.91 - 1.80)	60.00 - 1.90 (2.01 - 1.90)	60.0 - 1.92 (2.03 - 1.92)	60.0 - 2.10 (2.23 - 2.10)	60.00 - 1.74 (1.79 - 1.74)	30.0 - 1.81 (1.92 - 1.81)
Nr. observations	148324 (18433)	114899 (14881)	200217 (24572)	136161 (21687)	129380 (17654)	102785 (14554)	174136 (8286)	158322 (25502)
Nr. unique reflections	40742 (5910)	39139 (5682)	46863 (6983)	39808 (6335)	37993 (5760)	30243 (4721)	50839 (3496)	46163 (7353)
Completeness (%)	98.0 (88.6)	94.8 (85.7)	97.8 (91.9)	98.0 (97.3)	98.0 (92.9)	99.0 (96.5)	96.8 (91.1)	99.2 (98.8)
Multiplicity	3.6 (3.1)	2.9 (2.6)	4.3 (3.5)	3.5 (3.4)	3.4 (3.1)	3.4 (3.1)	3.4 (2.4)	3.4 (3.5)
$\langle I/\sigma(I)\rangle$	18.79 (2.37)	7.19 (1.93)	19.29 (3.67)	23.79 (4.92)	20.28 (3.74)	16.78 (4.47)	11.81 (3.06)	8.99 (1.92)
R-merge (%) [a]	4.4 (48.1)	13.0 (73.2)	4.5 (32.4)	3.1 (26.1)	3.7 (27.5)	5.1 (28.8)	5.9 (29.1)	7.9 (54.1)
R-meas (%) [b]	5.2 (57.9)	15.1 (72.2)	5.1 (37.9)	3.6 (31.0)	4.4 (33.1)	5.9 (34.8)	6.5 (37.1)	9.3 (63.8)
Wilson B ($Å^2$)	36.5	37.4	32.7	37.3	35.7	38.2	35.1	34.3

Values in parentheses refer to the highest resolution shell. [a] R-merge = merging R-factor, $(\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|) / (\sum_{hkl} \sum_i I(hkl)) \times 100\%$; [b] R-meas = multiplicity independent R-factor, $\sum_{hkl} [N_{hkl}/(N_{hkl} -1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i (hkl) \times 100\%$, where I is the observed intensity, $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections, and N_{hkl} is their multiplicity (*61*)

Table 3.5 Final refinement statistics.

	$[Gd^{3+}]$		apo-MpgP		$[Mg^{2+}$-MPG] crystal 1		$[Mg^{2+}$-MPG] crystal 2	
Resolution limits (Å)	37.56 - 2.50 (2.66 - 2.50)		39.51 - 1.90 (1.97 – 1.90)		20.85 - 1.90 (1.97 - 1.90)		42.16 - 1.80 (1.87-1.80)	
R-factor [a]	0.195 (0.228)		0.173 (0.242)		0.182 (0.293)		0.180 (0.215)	
nr. working reflections [b]	16926(2817)		38647 (3298)		36697 (3363)		44507 (3947)	
Free R-factor [c]	0.253 (0.286)		0.228 (0.303)		0.238 (0.317)		0.230 (0.273)	
nr. free reflections [b]	908 (140)		2073 (175)		1950 (193)		2342 (209)	
Overall coordinate error estimate (Å) [d]	0.34		0.28		0.24		0.24	
Model completeness and composition	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*
Model regions omitted	252-259	252-259	253-259	13-18, 254-259	251-259	255-259	251-259	255-259
Gd^{3+} ions	3	3	-	-	-	-	-	-
Cl^- ions	1	1	-	-	-	-	-	-
Mg^{2+} ions	-	-	-	-	-	1	1	1
MPG molecules	-	-	-	-	-	-	-	-
DMG molecules	-	-	-	-	-	1	1	1
$[PO_4]^{3-}$ / $[PO_3]^-$ ions	-	-	1	1	-	1	1	1
$[VO_3(O^{\delta 2}Asp6)]^{3-}$/ $[VO_3]^-$ ions	-	-	-	-	-	-	-	-
Non-hydrogen protein atoms	1940	1944	1960	1915	1939	1953	1956	1956
Disordered residues	0	2	2	1	1	1	4	2
Solvent molecules	146		439		323		391	

Mean B values (Å^2) [e]	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*
Protein main-chain	30.5	32.1	36.7	37.0	46.7	39.8	25.6	40.9
Protein side-chain	31.6	34.0	43.2	44.8	56.8	51.9	33.7	48.3
Ions and ligands	48.3	47.1	50.9	50.4	-	49	30.6	33.0
Solvent	28.3		43.3		55.5		43.5	
Model r.m.s. deviations from ideality								
Bond lengths (Å)	0.008		0.007		0.008		0.007	
Bond angles (º)	1.120		0.965		1.114		1.032	
Chiral centers (Å^3)	0.066		0.060		0.072		0.065	
Planar groups (Å)	0.005		0.005		0.005		0.005	
Model validation [f]								
% Ramachandran outliers	0.0		0.2		0.4		0	
% Ramachandran favored	95.6		98.0		95.4		97.6	
% Rotamer outliers	5.8		0.3		1.8		0.3	
C^{β} outliers	0		0		0		0	
Clash score	19.0		15.4		16.5		13.0	

[a] R-factor = $\sum_{hkl} ||F_o|-|F_c|| / \sum_{hkl} |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively;
[b] No $\sigma(F_o)$ cutoff ;
[c] Cross-validation R-factor computed from a randomly chosen subset of 5% of the total number of reflections, which were not used during the refinement;
[d] Maximum-likelihood estimate;
[e] Calculated from isotropic or equivalent isotropic B-values;
[f] Calculated with MolProbity (*41*)

Table 3.5 Final refinement statistics *(continued).*

	[Gd^{3+}-MPG] crystal 1		[Gd^{3+}-MPG] crystal 2		[MPG-VO^{2+}]		[Mg^{2+}-MG-HPO_4^{2-}] (Control)		[Mg^{2+}-MG-VO_4^{3-}] (Control)	
Resolution limits (Å)	24.96 - 1.90		28.79 -1.92		46.28 - 2.10		38.53 - 1.74		28.77 - 1.81	
	(1.97 - 1.90)		(1.99 - 1.92)		(2.17 - 2.10)		(1.80 - 1.74)		(1.87 - 1.81)	
R-factor [a]	0.162 (0.233)		0.161 (0.220)		0.156 (0.198)		0.172 (0.225)		0.168 (0.280)	
nr. reflections [b]	37877 (3648)		35831 (3271)		28710 (2700)		48242 (4569)		43834 (4274)	
Free R-factor [c]	0.203 (0.276)		0.214 (0.290)		0.211 (0.276)		0.219 (0.306)		0.213 (0.334)	
nr. reflections [b]	2020 (192)		1911 (167)		1521 (160)		2592 (253)		2340 (232)	
Overall coordinate error estimate (Å) [d]	0.27		0.25		0.29		0.24		0.25	
Model completeness and composition	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	Chain A	Chain B
Regions omitted	253-259	253-259	251-259	252-259	253-259	253-259	251-259	255-259	252-259	252-259
Gd^{3+} ions	-	-	-	-	-	-	-	-	-	-
Cl^- ions	1	-	-	-	-	-	-	-	-	-
Mg^{2+} ions	1	1	-	1	1	1	1	1	1	1
MPG molecules	-	-	-	-	-	-	-	-	-	-
DMG molecules	-	1	1	1	-	-	1	1	1	1
$[PO_4]^{3-}$ / $[PO_3]^-$ ions	-	1	-	-	-	-	1	1	-	-
$[VO_3(O^{\delta 2}Asp6)]^{3-}$/ $[VO_3]^-$ ions	-	-	-	-	-	1	-	-	1	1
Non-hydrogen protein atoms	1946	1946	1940	1946	1957	1954	1975	2020	1946	1959
Disordered residues	3	1	-	-	1	1	7	7	1	4
Solvent molecules	388		435		427		445		391	

Mean B values (Å^2) [e]	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	*Chain A*	*Chain B*	Chain A	Chain B
Protein main-chain	40.0	40.4	37.5	30.9	31.8	40.3	40.0	32.6	35.3	28.6
Protein side-chain	48.5	49.1	45.0	38.6	41.0	49.0	47.2	41.5	41.9	35.9
Ions and ligands	54.6	38.1	30.3	23.8	45.2	34.8	33.8	30.3	27.3	21.3
Solvent	49.7		46.8		46.7		47.9		44.2	
Model r.m.s. deviations from ideality										
Bond lengths (Å)	0.007		0.007		0.007		0.006		0.007	
Bond angles (º)	1.009		0.986		0.979		1.020		1.072	
Chiral centers (Å^3)	0.066		0.063		0.062		0.066		0.070	
Planar groups (Å)	0.005		0.005		0.005		0.005		0.005	
Model validation [f]										
% Ramachandran outliers	0		1.0		0.4		0.2		0	
% Ramachandran favored	97.6		96.2		97.4		96.7		97.2	
% Rotamer outliers	0.8		2.4		0.5		1.5		0.3	
C^{β} outliers	0		0		0		0		0	
Clash score	11.23		13.50		9.4		12.3		11.6	

[a] R-factor = $\sum_{hkl} ||F_o|-|F_c|| / \sum_{hkl} |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively;

[b] No $\sigma(F_o)$ cutoff ;

[c] Cross-validation R-factor computed from a randomly chosen subset of 5% of the total number of reflections, which were not used during the refinement.

[d] Maximum-likelihood estimate;

[e] Calculated from isotropic or equivalent isotropic B-values;

[f] Calculated with MolProbity (*41*).

3.6 REFERENCES

1. Empadinhas, N., Albuquerque, L., Henne, A., Santos, H., and da Costa, M. S. (2003) The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate, *Appl Environ Microbiol 69*, 3272-3279.
2. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, *J Mol Biol 361*, 1003-1034.
3. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E., and Kim, S. H. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states, *J Mol Biol 319*, 421-431.
4. Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2008) The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state, *Proc Natl Acad Sci U S A 105*, 5687-5692.
5. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction, *Science 299*, 2067-2071.
6. Pragai, Z., and Harwood, C. R. (2002) Regulatory interactions between the Pho and sigma(B)-dependent general stress regulons of *Bacillus subtilis*, *Microbiology 148*, 1593-1602.
7. Rao, K. N., Kumaran, D., Seetharaman, J., Bonanno, J. B., Burley, S. K., and Swaminathan, S. (2006) Crystal structure of trehalose-6-phosphate phosphatase–related protein: Biochemical and biological implications, *Protein Science 15*, 1735-1744.
8. Kawamura, T., Watanabe, N., and Tanaka, I. (2008) Structure of mannosyl-3-phosphoglycerate phosphatase from *Pyrococcus horikoshii*, *Acta Cryst D 64*, 1267-1276.

9. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) Large-scale identification of protein-protein interaction of *Escherichia coli* K-12, *Genome Res 16*, 686-691.
10. Aqvist, J., Kolmodin, K., Florian, J., and Warshel, A. (1999) Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data?, *Chem Biol 6*, R71-80.
11. Cleland, W. W., and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer, *Chem Rev 106*, 3252-3278.
12. Re, S., Imai, T., Jung, J., Ten-No, S., and Sugita, Y. (2011) Geometrically associative yet electronically dissociative character in the transition state of enzymatic reversible phosphorylation, *J Comput Chem 32*, 260-270.
13. Grzyska, P. K., Czyryca, P. G., Purcell, J., and Hengge, A. C. (2003) Transition state differences in hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions, *J Am Chem Soc 125*, 13106-13111.
14. Choe, J. Y., Iancu, C. V., Fromm, H. J., and Honzatko, R. B. (2003) Metaphosphate in the active site of fructose-1,6-bisphosphatase, *J Biol Chem 278*, 16015-16020.
15. Baxter, N. J., Bowler, M. W., Alizadeh, T., Cliff, M. J., Hounslow, A. M., Wu, B., Berkowitz, D. B., Williams, N. H., Blackburn, G. M., and Waltho, J. P. (2010) Atomic details of near-transition state conformers for enzyme phosphoryl transfer revealed by MgF_3^- rather than by phosphoranes, *Proc Natl Acad Sci U S A 107*, 4555-4560.
16. Cleland, W. W., and Hengge, A. C. (1995) Mechanisms of phosphoryl and acyl transfer, *FASEB J 9*, 1585-1594.
17. Webster, C. E. (2004) High-energy intermediate or stable transition state analogue: theoretical perspective of the active site and mechanism of beta-phosphoglucomutase, *J Am Chem Soc 126*, 6840-6841.

18. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem. 72*, 248-254.
19. Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc 2*, 2212-2221.
20. Gorrec, F. (2009) The MORPHEUS protein crystallization screen, *J Appl Cryst 42*, 1035-1042.
21. Borges, N., Marugg, J. D., Empadinhas, N., da Costa, M. S., and Santos, H. (2004) Specialized Roles of the Two Pathways for the Synthesis of Mannosylglycerate in Osmoadaptation and Thermoadaptation of *Rhodothermus marinus*, *J Biol Chem 279*, 9892-9898.
22. Kabsch, W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants, *J Appl Cryst 26*, 795-800.
23. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Cryst D 67*, 235-242.
24. Matthews, B. W. (1968) Solvent content of protein crystals, *J Mol Biol 33*, 491-497.
25. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs, *J Appl Cryst 37*, 843-844.
26. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Cryst D 58*, 1772-1779.
27. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE, *Z Kristallogr 217*, 644-650.

28. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated Structure Solution With autoSHARP, In *Macromolecular Crystallography Protocols* (Doublié, S., Ed), pp 215-230, Humana Press, Totowa, NJ.
29. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase, *Acta Cryst D 52*, 30-42.
30. Cowtan, K. (1994) 'dm': An automated procedure for phase improvement by density modification. , In *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, pp 34-38.
31. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.
32. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta Cryst D 62*, 1002-1011.
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst D 53*, 240-255.
34. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Cryst D 66*, 486-501.
35. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions, *Acta Cryst D 60*, 432-438.
36. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Cryst 66*, 213-221.
37. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion, *Acta Cryst 62*, 439-450.

38. Vagin, A. A., Murshudov, G. N., and Strokopytov, B. V. (1998) BLANC: the program suite for protein crystallography, *J Appl Crystallogr 31*, 98-102.
39. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Cryst D 59*, 1131-1137.
40. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) ARP/wARP and molecular replacement, *Acta Cryst 57*, 1445-1450.
41. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst D 66*, 12-21.
42. Velankar, S., and Kleywegt, G. J. (2011) The Protein Data Bank in Europe (PDBe): bringing structure to biology, *Acta Cryst 67*, 324-330.
43. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenke, l. A. (2008) Searching protein structure databases with DaliLite V.3., *Bioinformatics 24*, 2780-2781.
44. Krissinel, E., and Henrick, K. (2004) Secondary- structure matching (SSM), a new tool for fast protein alignment in three dimensions, *Acta Cryst. D60*, 2256-2268.
45. Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A. M., and Christendat, D. (2004) Structure- and Function-based Characterization of a New Phosphoglycolate Phosphatase from *Thermoplasma acidophilum*, *J Biol Chem 279*, 517-526.
46. Kabsch, W., and Sander, C. (1983) Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features, *Biopolymers 22*, 2577-2637.
47. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures, *J Appl Cryst 26*, 283-291.

48. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF - A program to identify and analyze structural motifs in proteins, *Protein Sci 5*, 212-220.
49. Westhead, D. R., Slidel, T. W., Flores, T. P., and Thornton, J. M. (1999) Protein structural topology: Automated analysis and diagrammatic representation, *Protein Sci 8*, 897-904.
50. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.
51. Ravelli, R. B., Leiros, H. K., Pan, B., Caffrey, M., and McSweeney, S. (2003) Specific radiation damage can be used to solve macromolecular crystal structures, *Structure 11*, 217-224.
52. Kabsch, W. (1976) Solution for Best Rotation to Relate 2 Sets of Vectors, *Acta Cryst A 32*, 922-923.
53. Hayward, S., and Berendsen, H. J. C. (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme, *Proteins: Struct, Funct and Genetics 30*, 144-154.
54. Empadinhas, N., and da Costa, M. S. (2010) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea, *Environ Microbiol* 13, 2056-2077.
55. Deng, H., Ray, W. J., Jr., Burgner, J. W., 2nd, and Callender, R. (1993) Comparison of vibrational frequencies of critical bonds in ground-state complexes and in a vanadate-based transition-state analog complex of muscle phosphoglucomutase. Mechanistic implications, *Biochemistry 32*, 12984-12992.
56. Cassano, A. G., Anderson, V. E., and Harris, M. E. (2004) Understanding the transition states of phosphodiester bond cleavage: Insights from heavy atom isotope effects, *Biopolymers 73*, 110-129.
57. Kirby, A. J., and Varvoglis, A. G. (1967) The Reactivity of Phosphate Esters. Monoester Hydrolysis, *J Am Chem Soc 89*, 415-423.
58. Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2005) HAD superfamily phosphotransferase substrate diversification: structure and function

analysis of HAD subclass IIB sugar phosphatase BT4131, *Biochemistry 44*, 8684-8696.

59. Koumanov, A., Ruterjans, H., and Karshikoff, A. (2002) Continuum electrostatic analysis of irregular ionization and proton allocation in proteins, *Proteins 46*, 85-96.
60. Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., and Allen, K. N. (2005) Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis, *Biochemistry 44*, 9404-9416.
61. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat Struct Biol 4*, 269-275.

3.7 ACKNOWLEDGEMENTS

The authors thank the European Synchrotron Radiation Facility (Grenoble, France) for support with the ID14-4 and ID23-1 data collections. This work was also carried out with the support of the Diamond Light Source (Didcot, U.K.), funded by the European Union Seventh Framework Programme grant nº 226716, and Fundação para a Ciência e Tecnologia (Portugal) grants PTDC/QUI/71142/2006, SFRH/BPD/29708/2006 to NB and SFRH/BD/23222/2005 to SG. The NMR spectrometers used are part of The National NMR Network of Portugal (REDE/1517/RMN/2005), supported by"Programa Operacional Ciência e Inovação (POCTI) 2010"and Fundação para a Ciência e a Tecnologia. Helpful suggestions from António Baptista (ITQB-UNL, Oeiras, Portugal) in the interpretation of the catalytic mechanism are also gratefully acknowledged.

CHAPTER 4

Reaction-coupled dynamics and insights into mechanistic aspects of the β-oxidative decarboxylation reaction by Isocitrate Dehydrogenase in *Escherichia coli* **

Direct evidence for the diad Tyr160-Lys230 in the assisting Brønsted acid-base catalysis*

Susana Gonçalves [1], Stephen P. Miller [2], Maria A. Carrondo [1], Anthony M. Dean [2] and Pedro M. Matias [1] †

[1] *Instituto de Tecnologia Química e Biologica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal*

[2] *Biotechnology Institute, University of Minnesota, 1479 Gortner Ave, St. Paul, MN 55108 USA*

** A manuscript has been recently submitted for publication

† Author´s contributions:

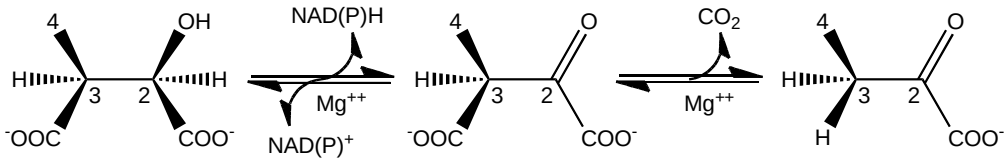
Prof. Maria Arménia Carrondo fostered the sabbatical leave of Prof. Antony M. Dean (A.M.D.) in Oeiras and promoted his collaboration with Dr. Pedro M. Matias (P.M.M) and Susana Gonçalves (S.G.) that led to the results presented in this chapter. She also participated in some initial discussions and provided funds to purchase some of the consumables used in this work. Cloning, expression and purification of wild type IDH and mutant IDH [K100M] was carried out by Stephen P. Miller under the supervision of A.M.D. Crystallization and crystal soacking assays, X-ray data collection, structural determination and refinement were carried out by S.G. under the supervision of P.M.M.

4.1 ABSTRACT

NAD(P)$^+$ dependent isocitrate dehydrogenase (IDH; EC 1.1.1.42) catalyses the oxidative decarboxylation of isocitrate to α-ketoglutarate in two steps, with dehydrogenation to the oxaloacetate intermediate preceding β-decarboxylation. Crystals of wild type *E. coli* IDH (*Eco*IDH) and its K100M variant were soaked with NADP$^+$, NADPH or thio-NADP$^+$ and either isocitrate or α-ketoglutarate in complex with Mg^{2+} or Ca^{2+}. The structures of a fully "closed" *pseudo*-Michaelis complex of the wild-type enzyme and a fully "closed" reaction product complex of the K100M variant are compared with the "quasi-closed" and "open" conformations to provide a comprehensive view of the induced fit needed for catalysis. Conserved catalytic residues binding the nicotinamide ring of the NADP$^+$ coenzyme and the metal-bound substrate, move as rigid bodies during domain closure by hinge motion that spans through the central β-sheet in the monomer. Key interactions established between the diad Thr105-Ser113 that flank the "phosphorylation loop", with the ribosyl-nicotinamide moiety, allows the productive binding of the cofactor at the active site pocket. Also, electrostatic interactions defined by the tetrad; Lys100-Leu103-Asn115-Glu336, play pivotal role for the assemblage of the catalytic competent active-site. Both the *pseudo*-Michaelis and the reaction product ternary complexes provided a complete picture of the mechanistic details of the dehydrogenation and β-decarboxylation steps. The structures are entirely in accord with the postulated roles for the Tyr160-Lys230* diad in Brönsted acid-base catalysis, as well as revealing the proton relay essential to catalysis.

4.2 INTRODUCTION

Escherichia coli isocitrate dehydrogenase [IDH; 2R,3S-isocitrate:$NADP^+$ oxidoreductase (decarboxylating); EC 1.1.1.42] belongs to a large superfamily of decarboxylating dehydrogenases evolutionarily unrelated to all other dehydrogenases like those that fall within into the D-lactate dehydrogenase-like class (*1, 2*). All family members use a two-step mechanism to catalyze dehydrogenations at the C2 and decarboxylations at the C3 of the 2R-malate core common to their substrates, viz:



The substrates differ at the 4 position: isocitrate (4 = $–CH_2COO^-$) is an intermediate in Krebs' cycle, isopropylmalate (4 = $–CH(CH_3)_2$) is an intermediate in leucine biosynthesis, homoisocitrate (4 = $–CH_2CH_2COO^-$) is an intermediate in the α-aminoadipate pathway for lysine biosynthesis in fungi (*3*) and tartrate (4 = –OH) is catabolized by *Pseudomonas putida* and *Agrobacterium vitis* (*4, 5*). Superpositioned active sites reveal functional residues interacting with the 2R-malate core are invariant, even among highly divergent family members (*6*).

Classic work on IDH catalysis (*7-13*) supports a two step mechanism, with C2 hydride transfer to the *re* face of the nicotinamide C4 preceding facile decarboxylation at C3 of the oxalosuccinate intermediate (Fig. 1.6). All known IDHs require a divalent a metal cation, such as Mg^{2+}, for catalysis. Structural studies of IDH confirm the expected bidentate chelation of Mg^{2+} by the carboxyl and hydroxyl at C1 of the substrate needed to stabilize the negative charges formed on the hydroxyl oxygen during the transition states of both steps (*7, 14*). IDH ternary complex structures also confirm that hydride transfer is to the *re* face of the nicotinamide C4 (*15-17*).

Less certain are the identities of the catalytic base, needed to initiate dehydrogenation by abstracting the proton from the isocitrate C2 hydroxyl, and of the catalytic acid, needed to protonate C3 following decarboxylation. The original suggestion (*14, 18*) that Asp283* [asterisk (*) denotes residues in the symmetry-related monomer] acts as the base, always dubious given that chelation to Mg^{2+} should reduce its pK_a far below that required for function under physiological conditions, failed to receive support upon detailed investigation (*19*). Structurally conserved water has also been proposed as the catalytic base (*20, 21*), with a proton relay system exiting the active site to allow exchange with bulk solvent, albeit in the absence of a specific mechanism to reduce its pK_a to within a plausible physiological range. Replacing the active site Tyr160 by phenylalanine compromised overall activity with isocitrate; decarboxylation of oxalosuccinate was unaffected but its reduction to isocitrate was severely lowered (*22*). These results support a role for Tyr160 in dehydrogenation, although in all binary complexes of IDHs (*20, 23-26*) and isopropylmalate dehydrogenase [IMDH, (*27*)] and in the ternary complexes of IDH (*21, 24, 28-33*) and tartrate dehydrogenase [TDH, (*34*)], the tyrosine side chain is not positioned to act as a catalytic base. Based on its geometry and spatial proximity to isocitrate in a binary complex, Lys230*, was proposed to be the proton donor following decarboxylation (*14*). Support for this conjecture was obtained upon replacing Lys230* by methionine which greatly compromised overall activity; decarboxylation of oxalosuccinate was obliterated while its reduction to isocitrate was barely affected (*22*). Detailed kinetic studies of mutations at the equivalent sites in pig heart IDH (Tyr140 and Lys212*) suggest Tyr160 is the acid that protonates the substrate after decarboxylation, with Lys230* playing a supporting role (*35*). Recent work on homoisocitrate dehydrogenase (*6*) has proposed that Lys199* and Tyr126 (equivalent to Lys230* and Tyr160 in *Eco*IDH) work as pair, with Lys199* abstracting the proton directly from the C2 hydroxyl during dehydrogenation to the oxaloacid, then reprotonating the hydroxyl during C3 decarboxylation to the

enol, and then once again abstracting the proton as Tyr126 protonates C3 to the ketone product.

The confusion surrounding the identities and exact roles of the active site residues is partly attributable to the absence of crystal structures detailing the active site during catalysis. A particular difficulty with *Eco*IDH (and by inference, other family members too) is that it undergoes a substantial conformational change upon substrate binding, not to a productive Michaelis complex, but rather to an intermediate nonproductive position. Structural data, though informative, may be misleading. Mutant kinetic data may also mislead when another active site residue partially compensates for loss of a functional group. Here, we determine the structures of ternary complexes of *Eco*IDH with bound substrate and coenzyme to better resolve the conformational changes needed to yield a productive Michaelis complex and reconcile conflicting hypotheses regarding the catalytic mechanism.

4.3 MATERIALS AND METHODS

Protein production and purification

E. coli strain ST∆*icd* (*36*).was transformed with plasmid (pIDH[wildtype] or pIDH[K100M] containing *icd* wildtype and Lys100Met mutant genes respectively) and grown at 37° to full density in 5 liters of broth (12 g tryptone, 24 g yeast extract, 4 g glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4 and 5 mg/L tetracycline) in a 6 L NBS BioFlow 3000 fermentor (University of Minnesota Biotechnology Resource Center). Cells were harvested by centrifugation and the paste flash frozen between blocks of dry ice and stored at -80° until needed.

Fifty g of cell paste were resuspended in 150 mls of extract buffer (10 mM KH_2P0_4, 0.5 M NaC1, 2 mM $MgCl_2$, and 2-mercaptoethanol, pH 7.7) and stirred for 1 hr at room temperature to allow IDH to become fully dephosphorylated. The suspension was then chilled to 4°, sonicated and the

cell debris removed by centrifugation. Enzymes were purified by a standard protocol (*37*) involving ammonium sulfate precipitation, DEAE (Pharmacia) anion chromatography and affinity chromatography using Affi-Gel Blue (BioRad). Protein concentrations were determined by the method of Bradford (*38*) using BSA as a standard. All preparations are 98% free of contaminating proteins as judged by Coomassie Blue staining following SDS PAGE.

Crystallization of wtIDH and IDH (K100M)

Concentrated protein solutions of *wt*IDH (35 *m*g/mL) and IDH (K100M) (45 *m*g/mL) were diluted to 20 *m*g/*m*L in a storage buffer solution containing 0.9 *m*M citric acid, 3.5 *m*M Na_2HPO_4 pH 6.0, 100 *m*M NaCl, 0.02% NaN_3 and 2 *m*M DTT. The crystal growth optimization was carried out using the hanging drop vapor diffusion method by screening 24 different crystallization solutions based on that established by Hurley and co-workers (*1*) by varying the pH and NH_4SO_4 concentration. The best crystallization solution for *wt*IDH was 1.85M NH_4SO_4, 50 *m*M citric acid/Na_2HPO_4 buffer pH 5.8, 0.1 M NaCl and 0.2 M DTT, whereas that for IDH (K100M) was 1.85M NH_4SO_4, 50 *m*M citric acid/Na_2HPO_4 buffer pH 5.2, 0.1 M NaCl and 0.2 M DTT. Protein crystals for ligand soaking and diffraction data collection were also obtained by the hanging drop vapor diffusion method. 2 μL drops were set up at 293 K in a XRL 24-well crystallization plate (Molecular Dimensions), by mixing native and mutant protein solutions with their respective crystallization solutions in a 1:1 ratio of protein to reservoir solution, and equilibrated against 500 μL of crystallization solution in the reservoir. Single crystals with tetragonal bipyramidal shape developed within 5 days, reaching dimensions of 500 μm x 200 μm x 100 μm for *wt*IDH and of 300 μm x 150 μm x 75 μm for IDH (K100M) (Fig. 4.1).

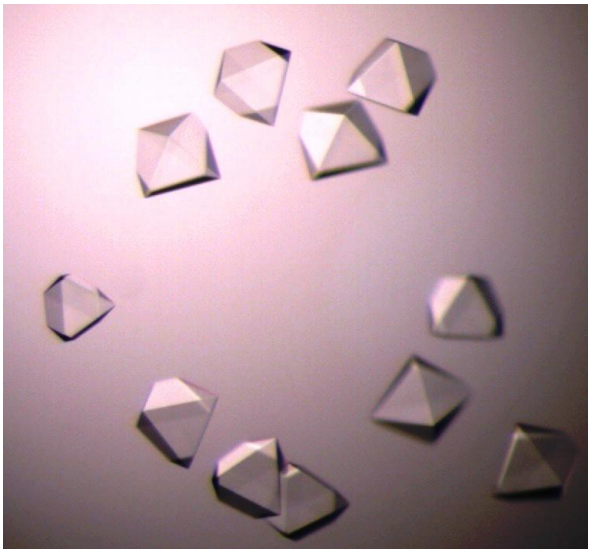


Figure 4.1 Tetragonal bipyramidal crystals of *Eco*IDH.

Crystal soaking procedure

Crystals of *wt*IDH and IDH (K100M) were transferred from their mother liquor into the final soaking solution in either one or two steps. In the two-step procedure the crystals were first soaked in Solution 1 (1.58 M NH_4SO_4, 156 *m*M NaHEPES pH 6.0, 52 *m*M Ca^{2+}/Mg^{2+}, 300 *m*M ICT) and then transferred to Solution 2 (1.58M NH_4SO_4, 156 *m*M NaHEPES/MES pH 7.5~8.0, 52 *m*M Ca^{2+}/Mg^{2+}, 300 *m*M ICT/α-KG, 500-400 *m*M $NADP^+$/NADPH/thio-$NADP^+$), whereas the one-step procedure involved only Solution 2. All the soaks were done at 293 K. Crystals were then cryo-protected by transferring them directly into an artificial mother liquor with the same composition as that of soaking solution 2 supplemented with 25% glycerol, and subsequently flash-cooled in N_2(l) prior to data collection. For each dataset, the composition of the soaking solutions and the final ligand concentrations, as well as the duration of each soak are described in Table 4.1.

Hereafter, the nomenclature used to designate the different datasets and the corresponding crystal structures is based on the different protein crystals and ligands in which they were soaked, using the following notation:

A [B:C:D], where A is either absent for *wt*IDH or K100M for IDH (K100M); B is either Mg^{2+} or Ca^{2+}; C is either ICT or α-KG; and D is either $NADP^+$, thio-$NADP^+$ or NADPH. In the one instance where two separate datasets were collected from different crystals of a given soak, the suffixes x1 and x2 are appended to distinguish between them.

The following stock solutions were used to prepare the soaking solutions: 1 M NaHEPES pH 7.5 (Sigma-Aldrich), 1 M MES-NaOH (Sigma-Aldrich) pH 6.3, 2 M $MgCl_2$ (Sigma-Aldrich); 2 M $CaCl_2$ (Sigma-Aldrich); 2 M ICT (Fluka) in 11% NH_4OH (Merck) pH 6.0; 0.5 M α-KG (Sigma-Aldrich) in 0.1 M MES-NaOH (Sigma-Aldrich) pH 6.3; and 0.24 M α-NADPH (Sigma-Aldrich) in 166 *m*M α-KG (Sigma-Aldrich), 166 *m*M MES-NaOH (Sigma-Aldrich) pH 6.3 and 1.54 M NH_4SO_4 (Merck). Na_2β-$NADP^+$ (Fluka) and Naβ-thio-$NADP^+$ (Sigma-Aldrich) were added directly to soaking solution 1 (see Table 4.1).

Data collection and processing

Datasets of the different IDH crystal soaks were collected at 100 K using different X-ray sources and detector systems. The [Ca^{2+}:ICT:$NADP^+$] and K100M [Ca^{2+}:α-KG:NADPH] datasets were measured *in-house*, using a Bruker-AXS Proteum Pt135 CCD detector system coupled to a Bruker-AXS Microstar-I rotating anode X-ray generator with Montel mirrors. The data were integrated with SAINT and scaled with SADABS as part of the Bruker-AXS Proteum Software Suite. Diffraction data statistics were obtained with XPREP (Bruker-AXS) and are listed in Table 4.2. The K100M [Mg^{2+}:ICT:$NADP^+$] x1 and K100M [Mg^{2+}:ICT:$NADP^+$] x2 datasets were collected at the ESRF (Grenoble, France) ID14-4 and ID23-1 beamlines respectively, using an ADSC Quantum Q315r detector. The K100M [Mg^{2+}:ICT:thio-$NADP^+$] and the [Ca^{2+}:ICT:thio-$NADP^+$] datasets were recorded at the SLS (Villigen, Switzerland) PXIII beamline, using a MAR225 detector. For the datasets collected at synchrotron X-ray sources, the diffraction images were processed with XDS (*39*), and the data collection statistics are included in Table 4.2. All diffraction data were further processed with the CCP4 Program Suite (*40*). Two preliminary

datasets, replicates of [Ca^{2+}:ICT:$NADP^{+}$] and K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 datasets, were initially collected *in-house* at 293K and the respective crystal structures were solved by Molecular Replacement with PHASER (*41*), using the coordinates of the PDB entry 1ai2 (*28*) as the search model. The solutions were partially refined (our unpublished results) and subsequently used as search models to solve the crystal structures herein described by the Molecular Replacement with PHASER. Although Matthews coefficient calculations (*42*) suggested the presence of 2 molecules in the asymmetric unit of all crystal structures, with V_m values of about 2.2 $Å^3Da^{-1}$ and a predicted solvent content of about 45%, the results of the Molecular Replacement calculations showed that, as in the structure reported by Hurley *et al.* (*1*), only one *wt*IDH or IDH (K100M) monomer was present in the asymmetric unit of each crystal structure, corresponding to V_m values and estimated solvent contents of *ca.* 4.5 $Å^3Da^{-1}$ and 72.4 % respectively.

Crystallographic refinement

The structures were refined using the amplitude-based Maximum-Likelihood target function with automatic weight optimization procedure as implemented through the graphics user interface of the PHENIX v.1.6.4 software package (*43*). The initial refinement step consisted of five macrocycles, the first of which comprised a rigid body refinement followed by individual coordinate and isotropic atomic displacement parameter (ADP) refinement. In the second and fourth macrocycles, a simulated annealing refinement (cartesian and torsion angles) was carried out using the default parameters, followed by individual coordinate and ADP refinement of individual isotropic B-factors. Monomer library descriptions for new ligands were created with LIBCHECK (*44*) through the smiles translator interface in Coot (*45*). Ligand restraints were further reviewed using JLigand (*46*). During refinement, the models were periodically inspected and corrected with Coot against sigmaA-weighted $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps. Inclusion of ordered solvent molecules was done with ARP/wARP (*47*), followed by

inspection in Coot. In the final refinement, automatic atom occupancy refinement was included in the protocol. The stereochemical quality of each model was assessed with MolProbity (*48*). A summary of the refinement statistics, model composition and stereochemical quality is presented in Table 4.2. Coordinates and structure factors were deposited with the PDBe (*49*) and the respective accession codes are included in Table 4.2.

Structure analysis

Structures were superposed at their corresponding domains I and II (see "Structural analysis" in the "Results" section for definitions of domains I and II), and r.m.s. deviations between superposed atoms were calculated using the CCP4 suite program LSQKAB (*50*). Domain rotations and hinge points were determined using the program DynDom (*51*) in the CCP4 suite (see "Interdomain hinge dynamics" in the "Results" section for details). The analysis of the residues involved in the binding of $NADP^+$/NADPH and ICT/α-KG (as a metal-bound ligand complex), was done by least-squares superposition of domains I and II respectively, using the CCP4 suite program LSQKAB. The calculation of the r.m.s. deviations between the superposed ligands was done by matching them using the least-squares superposition function implemented in Coot (*45*). Comparison of the IDH structures reported herein with their closest homologues was carried out using the Secondary Structure Matching (SSM) method (*52*) as implemented in the CCP4 suite, as an alternative to LSQKAB. Distances and angles were calculated using Coot. Crystal-packing contacts were determined with the CCP4 suite program NCONT. Most of the CCP4 suite programs were used through the CCP4 graphical user interface (*53*). Figures were generated using PyMOL Molecular Graphics System, Version 0.99rc (Schrödinger, LLC). Coordinates and structure factors were deposited in the Protein Data Bank in Europe (*49*). The accession codes for the coordinate entries and the respective structure factors are: 4aj3 and r4aj3sf for [Ca^{2+}:ICT:NADP], 4ajs and r4ajssf for K100M [Mg^{2+}:ICT:$NADP^+$] x1, 4ajr and r4ajrsf for K100M [Mg^{2+}:ICT:$NADP^+$] x2, 4ajc

and r4ajcsf for K100M [Ca^{2+}:α-KG:NADPH], 4ajb and r4ajbsf for K100M [Mg^{2+}:ICT:thio-$NADP^+$], 4aja and r4ajasf for [Ca^{2+}:ICT:thio-$NADP^+$]. The coordinate accession codes are also included in Table 4.3.

4.4 RESULTS

Structural analysis

Both wtIDH and IDH (K100M) crystallized in the tetragonal space group $P4_32_12$ with one monomer in the asymmetric unit (Fig. 4.2A). The biological unit is a homodimer with subunits related by a binary crystallographic axis (Fig. 4.2C). The monomers fold as previously described (*1*) with a large α+β domain (residues 1-124 and 318-416) and a small α/β domain (residues 125-317) that includes a distinctive clasp-like α/β sub-domain (residues 158-202). Secondary structure elements are referenced using Hurley's nomenclature (*1*) with residues in the second subunit denoted by an asterisk (*).

The small domain is similar in all crystal structures (0.27-0.34 Å C^α r.m.s. deviations), although deviations up to 2.0 Å are found in the loop linking β-strand "K" to the disordered β-strand "L" (residues 259-261) at the end of the sheet. The large domain is more variable (1.1 to 1.2 Å C^α r.m.s. deviations) with major structural differences found in the adenosine 2',5'-diphosphate binding pocket (residues 336-352 of the NADP-binding loop, residues 100-113 of the phosphorylation loop, residues 114-122 at the N-terminus of helix "d" and residues 71-79 that comprise helix "l"). Similar r.m.s. deviations are obtained when superpositioning domains from our [Ca^{2+}:ICT:NADP] *pseudo*-Michaelis complex on that of another Ca^{2+}-inhibited *pseudo*-Michaelis complex of *Eco*IDH (PDB 1ai2) (*28*) and on the apo-isoform of *Eco*IDH (PDB 1sjs) (*54*). Local structural differences attributable to crystal packing are present only at the interface defined by helix "h" and the phosphorylation loop of a second homodimer related by crystallographic symmetry. We thus attribute intramolecular structural rearrangements to the soaking assays.

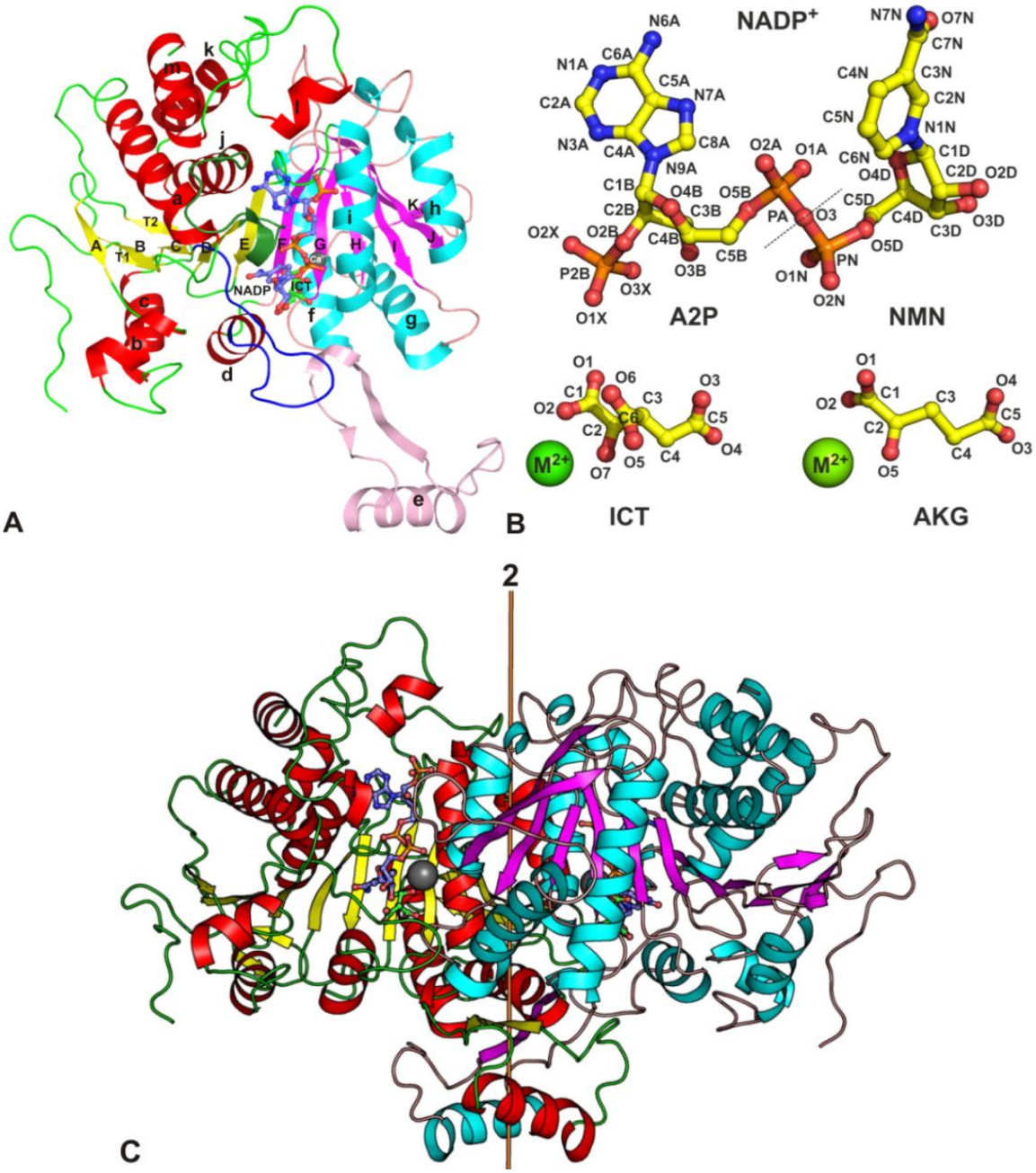

A
B
C
NADP+
A2P
NMN
ICT
AKG
NADP
ICT
2

Figure 4.2 The *E. coli* IDH structure and its ligands (A) Ribbon representation of the IDH monomer in the [Ca^{2+}:ICT:$NADP^{+}$] crystal structure, color coded according to domains I and II. Domain I (large $\alpha+\beta$ domain, residues 1-124 and 318-416) is represented by the red α-helices and yellow β-strands of the central β-sheet. Domain II comprises the small α/β region (residues 125-157 and 203-317) with α-helices in cyan and β-strands in purple and a α/β clasp-like region (residues 158-202) colored light pink. The secondary structure elements are labeled according to Hurley's scheme: β-strands are denoted by capital letters and α-helices are labeled in lowercase letters (*1*). The T1 and T2 elements indicate the N-terminal β-hairpin. The phosphorylation and NADP-binding loops are highlighted in blue and sky blue, respectively. The $NADP^{+}$ and ICT ligands are drawn in ball-and-stick and Ca^{2+} is depicted as sphere. Atom colors are: gray for calcium, orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms from $NADP^{+}$ and ICT are colored slate-blue and green, respectively. (**B**) Ball-and-stick diagrams of the $NADP^{+}$, ICT:M^{2+} and AKG:M^{2+} ligands in the IDH crystal structures showing the atom labeling scheme used in the text. M^{2+} represents either one of the Mg^{2+} or Ca^{2+} metal ions. In thio-$NADP^{+}$ the sulfur atom S1 replaces O7N. The dashed line at atom O3 in $NADP^{+}$ marks the division between the A2P and the NMN moieties. (**C**) The biological homodimer in the [Ca^{2+}:ICT:$NADP^{+}$] crystal structure. The two subunits are related by a crystallographic 2-fold symmetry axis (**2**, light brown) in space group $P4_32_12$. The first subunit (background monomer) is displayed with red α-helices and yellow β-strands and the second subunit (foreground monomer) is represented with α-helices in cyan and β-strands in purple. The $NADP^{+}$ and ICT ligands are drawn in ball-and-stick and Ca^{2+} is depicted as sphere. Atom colors are: gray for calcium, orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms from $NADP^{+}$ and ICT are colored slate-blue and green, respectively.

Interdomain hinge dynamics

Rigid body rotations occur about an axis at β-strands "E" and "F" between the large and small domains (Fig.4.3). Calculations were performed using the "open" apo-form of *Eco*IDH (PDB 1sjs) as a reference, aligning the small domains and then comparing the positions of the large domains, and hinge regions were equivalent to those determined by Finer-Moore and co-workers (*54*). Calculations with DynDom (*51*) showed that the large domain of the "closed" K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 ternary product complex had rotated 25°. DynDom calculations for the "closed" [Ca^{2+}:ICT:$NADP^{+}$] *pseudo*-Michaelis complex were less reliable and so we used LSQKAB (*50*) calculations to reveal that its large domain had rotated 24º into a similar position. The remaining structures displayed an intermediate conformation between the "open" and "closed" forms, with the large domains rotated 18-19º relative to the "open" apo-IDH structure (Fig.4.3). A second hinge axis crossing the α/β/α sandwich of the large domain on strands "C", "D" and "E" was revealed by comparing the relative orientations of the large domains of the "closed" K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 ternary product complex and its "*quasi*-closed" homologues. The relative positions of the large domains provide a static view of the intermediate conformational states that occur during activation of IDH upon substrate binding [1] (Fig. 4.4).

---

[1] In both tetragonal and orthorhombic crystal forms of *Eco*IDH the biological unit is a crystallographic dimer, i.e., the binary axis relating the two monomers coincides with a crystallographic 2-fold axis (Fig. 4.2C), and this precludes the observation of heterogeneous dynamics within the biological unit, since due to the effect of the crystallographic symmetry the hinge bending motion must be identical in both monomers, which may not be the case in solution.

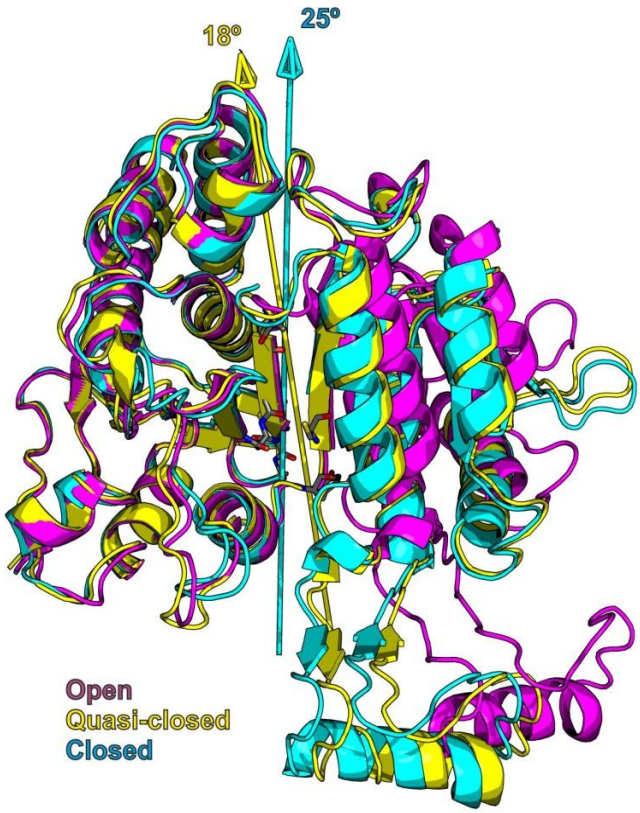


Figure 4.3 Interdomain hinge dynamics in *Eco*IDH. Representation of the C^{α} trace illustrating the rigid-body hinge motion of the domain II w.r.t. domain I. Hinge mechanical residues involved in the relative orientations of the domains were determined with DynDom (*51*) and are represented as sticks. IDH monomer conformations displayed are: "open" conformation in magenta (apo-IDH from PDB 1sjs, (*54*)), "quasi-closed" conformation in yellow (K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 structure) and "closed" conformation in cyan (K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 structure). The arrows indicate the hinge rotation axes and are colored according to the corresponding structure.

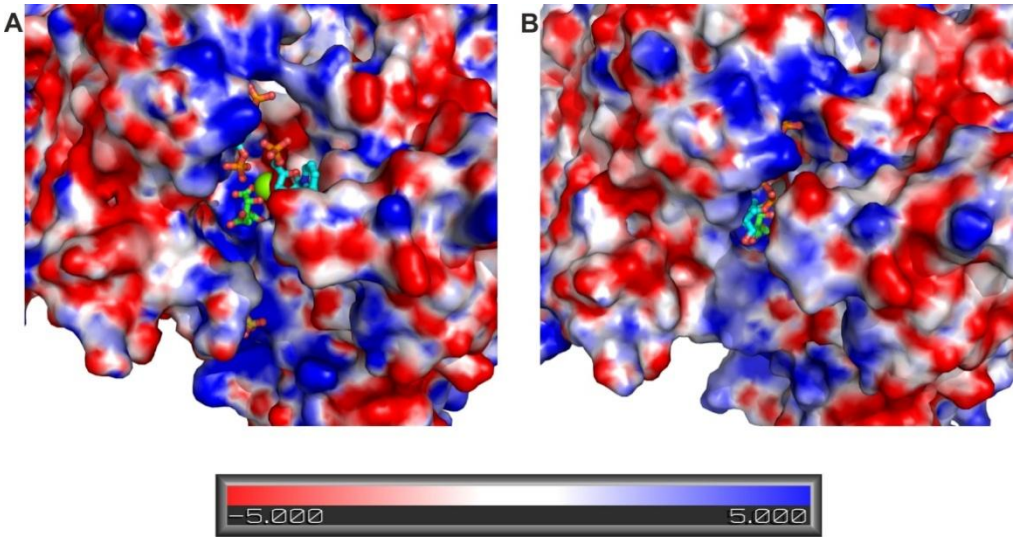


Figure 4.4 The catalytic pocket of *Eco*IDH (**A**) In the "quasi-closed" K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 structure (**B**) In the "closed" [Ca^{2+}:ICT:$NADP^{+}$] structure. The two structures were superposed at their domain II and illustrate the changes that take place in the catalytic pocket region with the hinge rotation of domain I from the "quasi-closed" to the "closed" conformation. Atomic charges and radii for the protein were calculated with the APBS v0.4.0 PyMol plug-in (*55*). The electrostatic distribution was determined by applying the non-linear Poisson-Boltzmann method, considering T = 300 K, an ionic strength of 0.15 M, a solvent dielectric constant ε_s = 80 and a protein dielectric constant ε_p = 6. The molecular surface was calculated using a solvent probe radius of 1.4 Å. The range of electrostatic potentials shown spans from -5 (red) to +5 *k*T/*e* (blue) units. The molecular ligands are shown in ball-and-stick representation and the Mg^{2+} and Ca^{2+} ions are drawn as spheres. Atom colors are: green for magnesium, gray for calcium, orange for phosphorus, blue for nitrogen, red for oxygen and green for carbon atoms in ICT; carbon atoms in $NADP^{+}$ and its degradation products are colored slate-blue.

$NADP^+$ and thio-$NADP^+$ Binding

Continuous electron densities were observed for the NADP(H) ligands in both “closed” [Ca^{2+}:ICT:$NADP^+$] and K100M [Mg^{2+}:ICT:$NADP^+$] x2 complexes (Fig. 4.5A and 4.5B). The isotropic atomic displacement parameters of the NADP(H) atoms were comparable to those of the side-chains interacting directly with the ligands. Bound NADP(H) adopts the same catalytically productive conformation in both structures (r.m.s. deviation of 0.14 Å for all ligand atoms). Electron densities corresponding to the adenosine 2',5'-diphosphate moiety (A2P), but not to the nicotinamide mononucleotide (NMN) moiety, of NADP were identified in the “quasi-closed” K100M [Mg2+:ICT:$NADP^+$] x1 and K100M [Ca^{2+}:α-KG:NADPH]) complexes (Fig. 4.5C) and refined with partial occupancies of 0.64 and 0.67, respectively. The isotropic atomic displacement parameters for the A2P atoms increased progressively from the adenine ring towards the 5´-phosphate, whereas those of the side-chains interacting directly with A2P remained approximately constant. Both ligands were bound similarly in the A2P binding pocket (r.m.s. deviation of 0.10 Å). Continuous well-defined electron densities with low isotropic atomic displacement parameters were observed for the thio-$NADP^+$ analogues in the [Ca^{2+}:ICT:thio-$NADP^+$] and K100M [Mg^{2+}:ICT:thio-$NADP^+$] complexes (Fig. 4.5D), with refined occupancies of 0.79 and 0.75, respectively. Bound thio-$NADP^+$ adopts the same catalytically unproductive conformation in both structures (r.m.s. deviation of 0.12 Å for all ligand atoms).

.

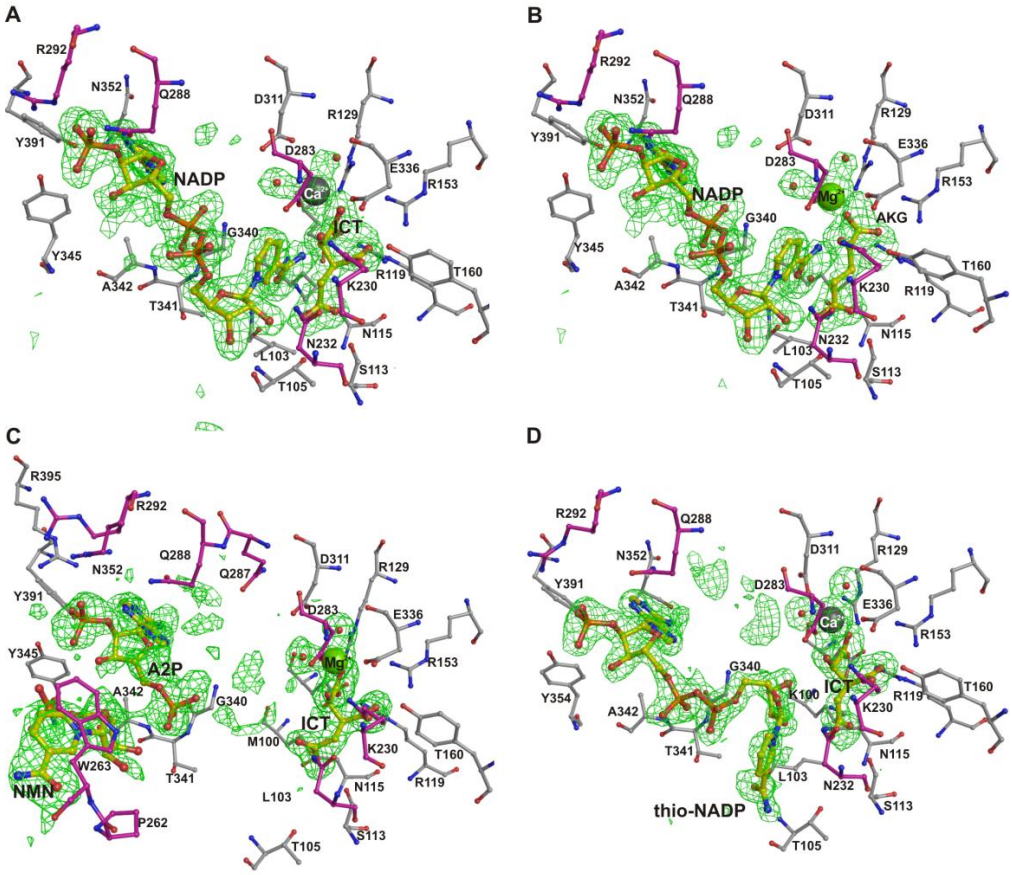


Figure 4.5 Ligands in the binding pocket of the *Eco*IDH crystal structures. (**A**) [Ca^{2+}:ICT:$NADP^{+}$]. (**B**) K100M [Mg^{2+}:ICT:$NADP^{+}$] x2. (**C**) K100M [Mg^{2+}:ICT:$NADP^{+}$] x1. (**D**) [Ca^{2+}:ICT:thio-$NADP^{+}$]. Unbiased sigmaA-weighted $|F_o|$-$|F_c|$ electron density maps covering the ligands in the catalytic pockets were obtained with PHENIX by simulated annealing in the absence of the ligands and water molecules located inside the catalytic pocket and coordinated either to the cocatalytic Mg^{2+} metal ion or to its analogue Ca^{2+}. The maps are represented as a green mesh and contoured at the 3σ level for all structures except for K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 structure (**C**) where it is drawn at the 3σ level for the [Mg^{2+}:ICT] complex and at the 2σ level for the hydrolyzed $NADP^{+}$. Protein residues and polyatomic ligands are drawn in ball-

and-stick representation, and the Mg^{2+} and Ca^{2+} metal ions and the water molecules are drawn as spheres. Atom colors are green for magnesium, gray for calcium, orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms are colored yellow in the ligands, and light-gray and magenta respectively in the first and second subunits of the biological homodimer.

The adenosine-2´,5´-biphosphate (A2P) binding site

The A2P binding site is defined by the interdomain 3/10 helix (residues 318-324) on one side of the adenine ring, the NADP-binding loop (residues 336-352) on the opposite side, and α-helix "l" (residues 390-397) which is oriented obliquely to the plane of the ring (Fig. 4.6). In all structures the adenine ring π-stacks with His339, the adenine ring and the endocyclic ribose oxygen adopt the typical *anti*-conformation with respect to each other around the N-glycosidic bond, and the ribose ring binds in an approximate half-chair C2'-exo/C3'-endo conformation (Fig. 4.7A) (*21, 28, 33, 56*). The nitrogen atoms of the adenine ring are involved in several hydrogen bonds: N6A to the carbonyl oxygen of Asn352, N1A to the amide nitrogen of Asn352, and N3A to water W10. This water forms part of a hydrogen-bonded chain of five solvent molecules (W10-W14), present only in the "closed" conformation of IDH (Fig. 4.7A), linking α-helices "i" and "h*" to α-helix "l" of the large domain. The protein atoms involved in this network are Asp392 $O^{\delta 2}$, Gln287* $N^{\varepsilon 2}$, Arg395 $N^{\eta 2}$, and Asp392 O.

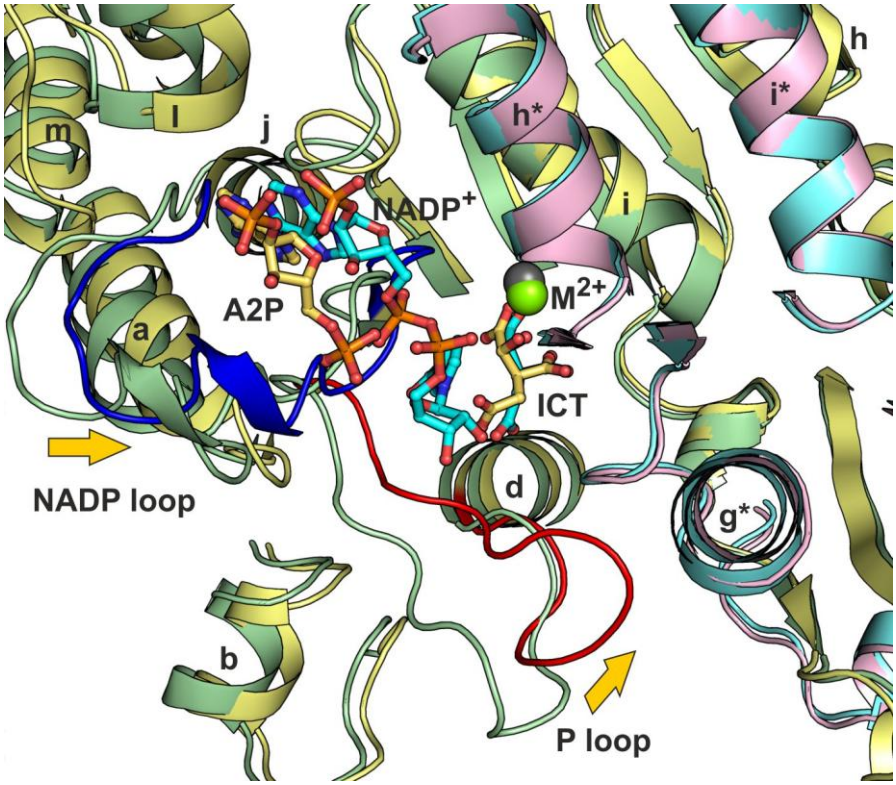


Figure 4.6 Bound ligands in the catalytic pocket of the "closed" and "*quasi*-closed" structures of *Eco*IDH. Two homodimers representing the "closed" and "*quasi*-closed" structures *of Eco*IDH were superpositioned at their C^{α} carbon atoms using the CCP4 Superpose program (*40*) and are drawn as ribbon diagrams. The "closed" structure is represented by the [Ca^{2+}:ICT:$NADP^{+}$] homodimer (the first subunit is colored light yellow and the second subunit is drawn in cyan) and the "*quasi*-closed" structure by the K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 homodimer (the first subunit is displayed in light green and the second subunit is represented in light pink). In the "closed" conformation, the NADP-binding loop (NADP loop) and the phosphorylation loop (P loop) are highlighted in blue and red colors respectively. Arrows indicate the concerted loop motions that accompany the enzyme transition from the "*quasi*-closed" to the fully "closed" and active state. The A2P, $NADP^{+}$ and ICT ligands are shown in ball-and-stick representation, and the Mg^{2+} and

Ca^{2+} ions as spheres. Atom colors are gray for calcium, green for magnesium, orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms are colored light yellow and cyan for the hydrolyzed (A2P) and intact forms of $NADP^+$, respectively.

In all structures O1X and O3X of the 5´-phosphate group hydrogen bond to the side chain hydroxyls of Tyr345 and Tyr391, respectively. In the "closed" conformations additional hydrogen bonds form between O2X and Gln288* $N^{\varepsilon 2}$ and Arg292* N^{ε}, O3X and Arg292* $N^{\eta 1}$, while W15 bridges O1X to Lys344 N and W10 bridges O2X to Gln287* $O^{\varepsilon 1}$ (Fig. 4.7A). The PA centre of the pyrophosphate hydrogen-bonds via O1A and O2A to the backbone amide of residues Gly340 and Ala342 in the NADP-binding loop (Fig. 4.7A). The NADP(H) pyrophosphate backbone, which adopts a PA-O3-PN-O5D dihedral angle of 92º, contributes to the assembly of a six atom ring formed by four water molecules, the C2-hydroxyl of isocitrate and the active site Mg^{2+} or Ca^{2+}.

In the "quasi-closed" conformations the A2P binding site is displaced as a rigid-body, with interactions between the large domain and α-helix "h*" replaced by bulk solvent molecules and by Arg395 $N^{\eta 1}$. The interaction with Gly340 is absent in structures with bound thio-$NADP^+$, and the pyrophosphate backbone adopts a different conformation with PA-O3-PN-O5D dihedral angles of 133º in the K100M [Mg^{2+}:ICT:thio-$NADP^+$] complex and 114º in the [Ca^{2+}:ICT:thio-$NADP^+$] complex. These places the thio-nicotinamide mononucleotide at the entrance of the isocitrate binding pocket in a conformation flipped relative to NMN in the "closed" structures (Fig. 4.5D).

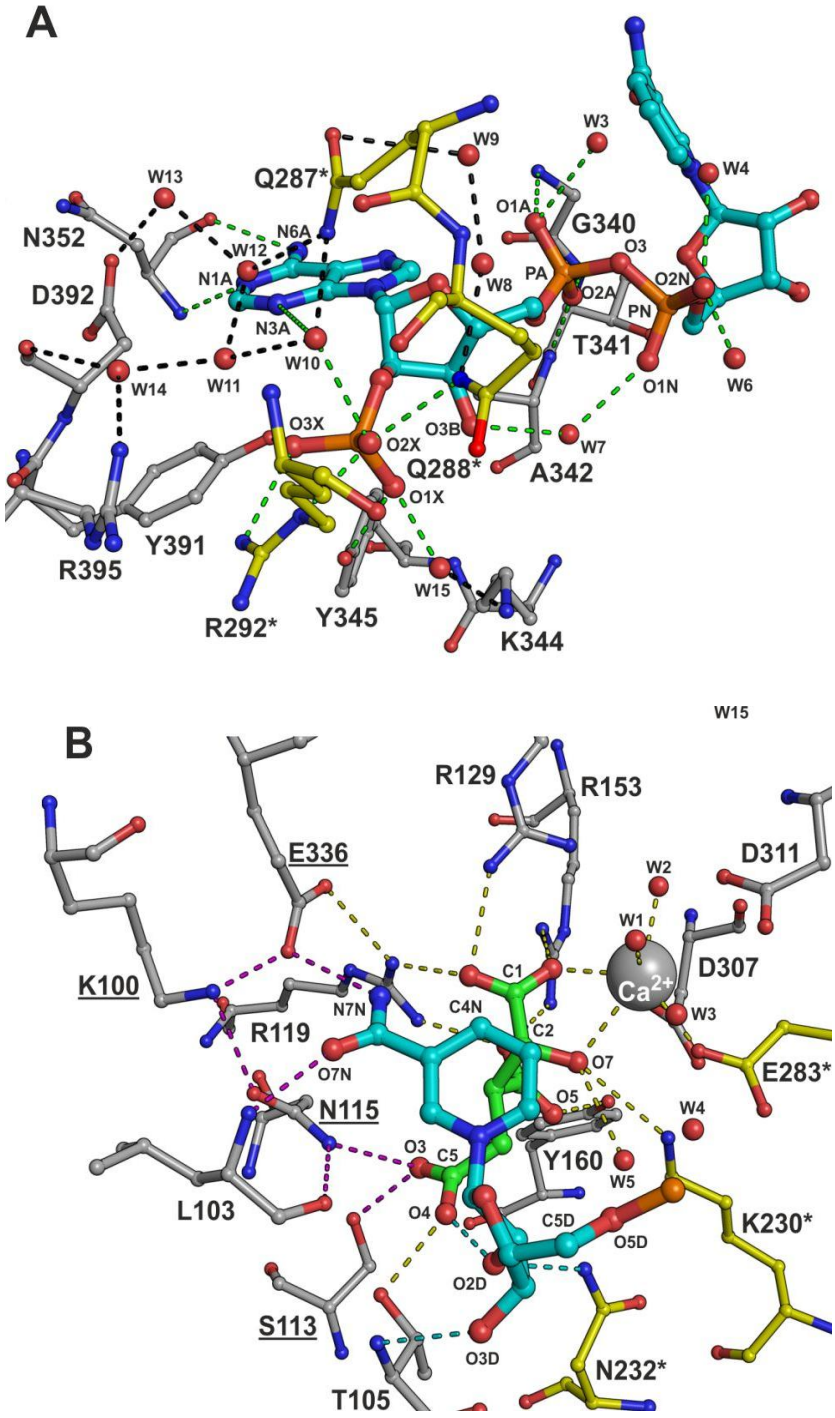

A
W13
Q287*
W9
W3
W4
O1A
G340
O3
N352
N6A
W12
N1A
PA
O2N
D392
W8
O2A
PN
N3A
T341
W10
W14
W11
O1N
W6
O3X
O3B
W7
O2X
Q288*
A342
O1X
Y391
R395
W15
R292*
Y345
K344
B
W15
R129
R153
E336
D311
W2
W1
K100
C1
D307
Ca2+
N7N
C4N
W3
R119
C2
O7
E283*
O7N
O5
W4
N115
O3
C5
Y160
W5
L103
O4
C5D
K230*
O5D
O2D
S113
O3D
N232*
T105

Figure 4.7. The NADP binding pocket. (**A**) The NADP adenosine-2´,5´-biphosphate (A2P) binding site. (**B**) The NADP ribosylnicotinamide-5´-phosphate (NMN) and [Ca^{2+}/Mg^{2+}:ICT] binding sites. Residues involved in the binding of the ternary complex of the "closed" [Ca^{2+}:ICT:$NADP^{+}$] structure and the NADP and ICT ligands are represented in ball-and-stick and the water molecules and the Ca^{2+} ion are shown as spheres. Dashed lines indicate hydrogen or coordination bonds and are color-coded according to the intervening functional groups of NADP, and to the substrate complex [Ca^{2+}:ICT]: green, for the interactions with the A2P moiety; cyan, for the ribosyl group of the NMN moiety; yellow for the interaction with the [Ca^{2+}:ICT] complex; black, for the interaction between the protein residues and the structural water molecules, and purple for the interactions involved in the anchoring of the nicotinamide moiety and assembly of the catalytic tetrad Ser113-Asn115-Lys100-Glu336 (highlighted with underlined residue labels). Carbon atoms are colored cyan for NADP, green for ICT, and light gray and yellow for protein carbon atoms of the first and second subunits of *Eco*IDH, respectively. The labels of the protein residues from the second subunit are suffixed with an asterisk and the ligand atoms are labeled according to Fig. 4.2B.

The nicotinamide mononucleotide (NMN) binding site

The NMN binding site lies adjacent to the isocitrate binding pocket (Fig. 4.7B). In both "closed" [Ca^{2+}:ICT:$NADP^{+}$] and K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 complexes, NADP(H) binds in a bent conformation with the nicotinamide ring above the substrate binding pocket, its amide oriented towards the large domain. The ribosyl moiety is anchored to the phosphorylation loop, in a roughly half-chair C2'-exo/C3'-endo conformation, with a dihedral angle of 166º about the O5D-C5D phosphoester bond. The 3´-hydroxyl O3D hydrogen-bonds to Thr105 N and Thr105 O. The 2´-hydroxyl O2D interacts with a C5-carboxylate oxygen of isocitrate or α-ketoglutarate, as well as with Asn232* $N^{\delta 2}$. These interactions trigger a shift in the phosphorylation loop towards the

active site pocket where it acts as a platform to bind NADP in a catalytically productive conformation (Fig. 4.6).

In the [Ca^{2+}:ICT:$NADP^{+}$] pseudo-Michaelis complex, the amide N7N hydrogen bonds to the C1-carboxylate of isocitrate (3.1 Å) and Glu336 $O^{\varepsilon 1}$ (2.9 Å), while the O7N hydrogen bonds with Leu103 N (2.9 Å) and Asn115 $N^{\delta 2}$ (3.0 Å) (Fig. 4.7B). Similarly, in K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 the N7N is within van der Waals distance of the C1-carboxylate of α-ketoglutarate (3.8 Å) and hydrogen bonds to Glu336 $O^{\varepsilon 1}$ (3.2 Å), while the O7N hydrogen bonds with water W66 (2.9 Å, and near Lys100 N^{ζ} in the *pseudo*-Michaelis complex) and Asn115 $N^{\delta 1}$ (2.9 Å) (Fig. 4.8A). The Lys100-Leu103-Asn115-Glu336 tetrad (Fig. 4.7B) plays a central role both in anchoring the nicotinamide ring and in triggering a cascade of electrostatic interactions that propagate towards the hinge region of the central β-sheet, facilitating the full closure of the enzyme. In particular, Asn115 plays a pivotal role in anchoring both the NMN moiety and isocitrate in a conformation favorable to catalysis.

In both "closed" complexes, the nicotinamide ring lies with its C4N poised to receive the hydride from isocitrate C2 on its *re* face (Fig. 4.7B and Fig. 4.8A). In structures with bound thio-$NADP^{+}$, the entire thio-NMN moiety is rotated 166º about the O5D-C5D phosphoester bond. The C4N now points away from the active site while the *si* face of the nicotinamide ring lies against the C5-carboxylate of isocitrate. This flipped conformation is stabilized by a hydrogen bond from the N7N of the thioacetamyl to Thr105 O and interactions between the ribosyl-moiety of thio-NMN and waters that fill the NMN binding pocket (Fig. 4.8B).

Evidence for the hydrolysed NMN moiety is given by its observation in the "closed" K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 and the "quasi-closed" K100M [Mg^{2+}:ICT:$NADP^{+}$] x1 complexes (Fig. 4.5C). The *si* face of the nicotinamide ring stacks against the aromatic ring of Trp263* with the carboxyamide stabilized by the loop formed by residues 258-261. Preliminary evidence for side-reaction hydrolytic activity in IDH was given by mass spectrometry analysis (unpublished results). These studies detected the presence of

hydrolysed $NADP^+$ products in re-dissolved tetragonal crystals of IDH soaked with $NADP^+$, but not when $NADP^+$ was added to the protein solution. Despite their qualitative nature, these results suggest that the cofactor hydrolysis should occur only in the “closed” or “quasi-closed” conformation of the enzyme. Further studies are required to establish the biological relevance of coenzyme hydrolysis by IDH.

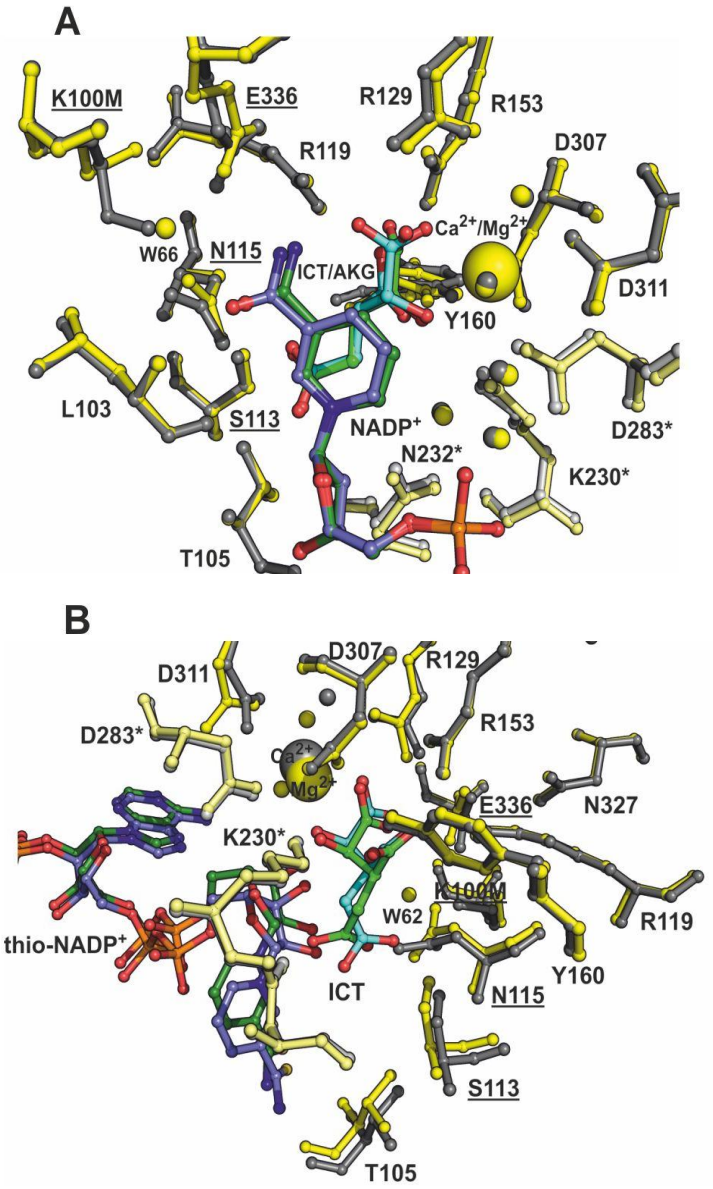


Figure 4.8 Structural comparison of the ternary complexes at the NMN and [Mg^{2+}:ICT] binding sites. (**A**) Comparison between the *pseudo*-Michaelis ternary complex ([Ca^{2+}:ICT:$NADP^+$] in gray) and product-bound complex (K100M [Mg^{2+}:ICT:$NADP^+$] x2 in yellow) structures of *Eco*IDH. In the *pseudo*-Michaelis ternary complex the carbon atoms of the NADP and ICT ligands are colored slate blue and cyan, respectively. In the product-bound complex, the carbon atoms of the NADPH and α-KG ligands are colored forest green and light green, respectively. (**B**) Comparison between the inactive ternary complexes in the [Ca^{2+}:ICT:thio-$NADP^+$] (gray) and K100M [Mg^{2+}:ICT:thio-$NADP^+$] (yellow) Structures. In the wild-type structure the carbon atoms of the thio-NADP and ICT ligands are colored slate blue and cyan, respectively, and in the mutant structure the corresponding colors are forest green and light green. Protein residues and polyatomic ligands are drawn in ball-and-stick representation and the Mg^{2+}, Ca^{2+} metal ions and the water molecules are drawn as spheres and colored according to their respective structure. The residues belonging to the second subunit in the biological dimer are in a lighter hue of the corresponding color used for the first subunit, and the respective residue labels are suffixed with an asterisk. The residues in the catalytic tetrad are highlighted with underlined residue labels.

The substrate binding pocket

As previously described (*14*), isocitrate and α-ketoglutarate bind in a pocket located at the base of the cleft between hinge axis strands “E” and “F” and helices “i” and “h*” at the dimer interface (Fig. 4.2A). In all structures a divalent metal ion, Mg^{2+} or Ca^{2+}, chelates a C1-carboxylate oxygen and the C2-oxygen of bound isocitrate or α-ketoglutarate (a hydroxyl and a carbonyl, respectively. The conserved Arg119, Arg129 and Arg153 provide a binding scaffold for the C1-carboxylate (Fig. 4.9) - this despite the 90º rotation around the C1-C2 σ-bond caused by the change in hybridization at C2 from sp^3 in isocitrate to sp^2 in α-ketoglutarate (Fig. 4.8A). A Lys230*-Tyr160 diad also

participates in the binding the C1-carboxylate and C2-hydroxyl of isocitrate, respectively (*22*). As with the NADP-binding pocket, the partial displacement of the large domain during hinge closure produces a positional shift in the substrate binding site, with preserved hydrogen bonds to the ligand C1 and C2 carboxylates.

Both isocitrate and α-ketoglutarate bind in "active" and "inactive" conformations, as reflected in the different C2-C3-C4-C5 dihedral angles: 139º ("active") and 63º ("inactive") for isocitrate (Fig 4.8B), 172º ("active") and 75º ("inactive") for α-ketoglutarate. In the "active" conformation, observed in the "closed" [Ca^{2+}:ICT:$NADP^+$] *pseudo*-Michaelis complex and K100M [Mg^{2+}:ICT:$NADP^+$] x2 ternary product complex, and also in the "quasi-closed" [Ca^{2+}:ICT:thio-$NADP^+$] complex, the C5-carboxylate hydrogen bonds with Ser113 O^{γ}, Thr105 O^{γ} and Asn115 $N^{\delta 1}$ (Fig. 4.9). In the inactive conformation, observed in the "quasi-closed" K100M [Mg^{2+}:ICT:$NADP^+$] x1, K100M [Mg^{2+}:ICT:thio-$NADP^+$] and K100M [Ca^{2+}:α-KG:NADPH] complexes, the C5-carboxylate hydrogen bonds to Ser113 O^{γ} and one or more waters.

The coordination number (CN) of Mg^{2+} changes from CN = 5 to CN = 6 as the enzyme shifts from the "quasi-closed" K100M [Mg^{2+}:ICT:$NADP^+$] x1 and K100M [Mg^{2+}:ICT:thio-$NADP^+$] to the "closed" K100M [Mg^{2+}:ICT:$NADP^+$] x2 positions. With Ca^{2+}, CN = 7 and CN = 6 in the "quasi-closed" [Ca^{2+}:ICT:thio-$NADP^+$] and K100M [Ca^{2+}:α-KG:NADPH] complexes respectively, whereas CN = 6 in the "closed" [Ca^{2+}:ICT:$NADP^+$] complex. The Ca^{2+} in the [Ca^{2+}:ICT:$NADP^+$] structure lies close to the Mg^{2+} in K100M [Mg^{2+}:ICT:$NADP^+$] x2 structure, effectively mirroring the Michaelis-Menten complex (Fig. 4.8A). In addition to binding the isocitrate or the α-ketoglutarate ligands, Mg^{2+} coordinates to Asp307 $O^{\delta 1}$ and two water molecules, W1 and W2, the second a part of a conserved six-atom ring involved in a proton relay mechanism. This non-labile pentavalent coordination sphere is supplemented with a sixth ligand, Asp283* $O^{\delta 2}$, that completes the Mg^{2+} coordination sphere in the K100M [Mg^{2+}:ICT:$NADP^+$] x2 and [Ca^{2+}:ICT:$NADP^+$] complexes. The Ca^{2+} coordination

sphere also includes a transient ligand, Asp311 $O^{\delta 2}$, coordinated to the metal ion prior to full closure of the enzyme (Fig 4.8B).

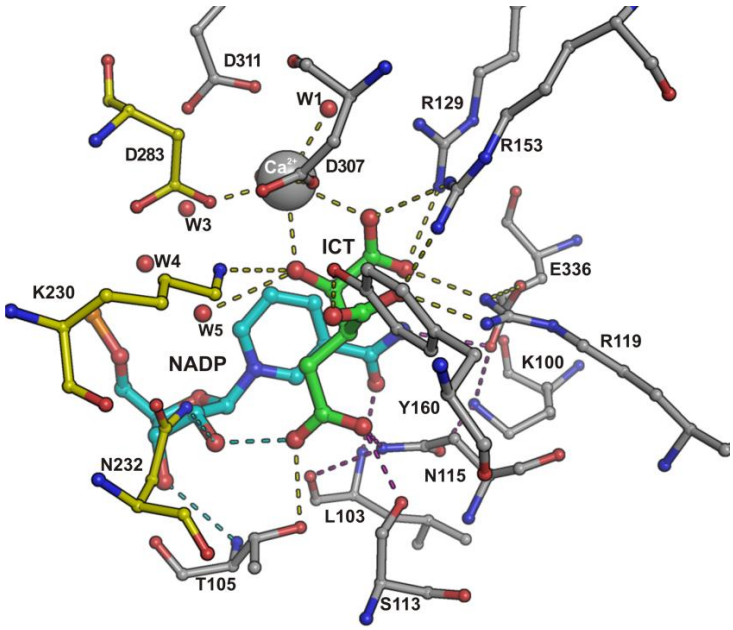


Figure 4.9 The substrate binding pocket. Representation of the [M^{2+}:ICT] binding pocket in the *pseudo*-Michaelis complex structure [Ca^{2+}:ICT:$NADP^{+}$]. The protein residues and the NADP and ICT ligands are represented in ball-and-stick, and the water molecules and the Ca^{2+} ion are shown as spheres. Dashed lines indicate hydrogen bonds between protein residues, water molecules and the ligands, and are color-coded according to the functional groups of NADP and the [Ca^{2+}: ICT] complex substrate: cyan, for the ribosyl group of the NMN moiety; yellow for the interaction with the [Ca^{2+}:ICT] complex, and purple for the interactions involved in the anchoring of the nicotinamide moiety and assembly of the catalytic tetrad Ser113-Asn115-Lys100-Glu336. Atom colors are: gray for calcium; orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms are colored cyan for NADP, green for ICT, and light gray and yellow for protein carbon atoms of the first and second subunits of *Eco*IDH, respectively.

4.6 DISCUSSION

Hinge dynamics and substrate-mediated structural rearrangements

Metal-dependent NAD(P)-linked hydroxyacid oxidative decarboxylases undergo structural rearrangements typical of many enzymes, including rigid body motions and substrate-induced conformational changes that modulate their catalytic activities (*57*). Studies of EcoIDH have identified fundamental principles behind substrate-induced conformational changes (*22, 24, 30, 54*). However, crystallographic intermediate-trapping methods based on crystal soaks with substrates and substrate analogues have proved insufficient to identify all the residues involved, mostly because crystal packing is incompatible with the fast structural rearrangements taking place during catalysis.

Our cryo-trapped snapshots, complexes [Ca^{2+}:ICT:$NADP^{+}$] and K100M [Mg^{2+}:ICT:$NADP^{+}$] x2, provide deeper insights into the structural conformations needed to produce productive ternary complexes. Both *wt*IDH and its K100M variant undergo a complete hinge closure that differs from the “quasi-closed” conformations previously reported. DynDom calculations provide a simple but effective overview of enzyme closure by comparing the orientation of the large and small domains in the “closed” and “quasi-closed” conformations relative to the “open” conformation (PDB 1sjs) (*54*). Pivotal residues are located throughout the central β-sheet spanning the large domain. We identified two distinct hinge regions at β-strands “C”, “D”, “E” in the “quasi-closed” state, and “E”, “F” in the “closed” state. These hinge motions trigger a cascade of electrostatic forces that propagate in concert with substrate binding. In the “quasi-closed” conformation, ligands bind at the pocket surface yielding non-productive complexes. In the “closed” conformation they move as rigid bodies to the inner pocket region along with a rotation of the large domain towards the small domain.

Pseudo-productive conformations have been reported for *Aeropyrum pernix* IDH (*Aper*IDH; PDB 1xkd, chain B) (*21*) and for *Eco*IDH mutant S113E

with bound isopropylmalate (PDB 1hj6) (*30*). In addition, cryo-trapped ternary complexes with bound NADPH, α-ketoglutarate and Ca^{2+} have been reported for *Saccharomyces cerevesiae* IDH isoform 1 (*Scer*IDH1, PDB 2qfx) (*31*) and the R132H mutant of human IDH isoform 1 (R132H *Hsap*IDH1; PDB 3inm) (*32*) Despite different electrostatic environments at their active sites, all these structures display features in common with the phosphorylation loops in our [Ca^{2+}:ICT:$NADP^+$] and K100M [Mg^{2+}:ICT:$NADP^+$] x2 complexes. In all cases, Thr105 and Ser113 (*Eco*IDH numbering) are conserved in the homologous enzymes and play pivotal roles in anchoring the C5-carboxylate group of isocitrate or α-ketoglutarate and the 3'-hydroxyl group O3D of the NMN. The major structural difference between the *pseudo*-productive conformations and our fully "closed" conformation is the shift in the phosphorylation loop towards the active site pocket which blocks access to bulk solvent while stabilizing the substrates in position for catalysis (Fig. 4.6). This loop shift also prevents hydrolysis of the coenzyme by *Eco*IDH. Thus, the significance of the extent of the phosphorylation loop structural rearrangement in *E. coli* and in its counterparts in other organisms extends beyond its regulatory role in the "on-off switch" mechanism (*54, 58-60*), by directly intervening in the assembly of a productive ternary complex, and in the product release during the catalytic turnover.

In addition to the Thr105-Ser113, the tetrad comprising Lys100-Leu103-Asn115-Glu336 is equally significant for the enzyme closure and for the productive anchoring of the nicotinamide ring. In particular, as seen in the "closed" [Ca^{2+}:ICT:$NADP^+$] and K100M [Mg^{2+}:ICT:$NADP^+$] x2 complexes, Asn115 plays a central role in anchoring the C5-carboxylate of isocitrate and α-ketoglutarate, as well as helping orientate the carboxamide of the nicotinamide ring (Fig. 4.7B).

Interestingly, in the K100M [Mg^{2+}:ICT:$NADP^+$] x2 structure, the now absent electrostatic contribution of the hydrogen bond between Lys100 and Asp336, which in turn anchors the nicotinamide ring in *wt*IDH, is rescued by a water molecule (W66) which is almost coincident with the position of the

Lys100 N^{ζ} atom (Fig. 4.8A). Preliminary catalytic rescuing assays showed that NH_4^+ activates the K100M mutant, perhaps by replacing the missing amine. Since IDH was crystallized using NH_4SO_4 as a precipitating agent, these results as well as the local negatively charged environment provided by O7N of the NMN moiety and the carboxylate group of Glu336 suggest that W66 may actually be an NH_4^+ cation, which is isoelectronic with H_2O and therefore indistinguishable by X-ray crystallography at 2.7 Å resolution.

The active site

Replacing Mg^{2+} by Ca^{2+} reduces k_{cat} > 2500-fold (*28*). Theoretical studies based on small distance and angle perturbations of the C2-hydroxyl group of isocitrate and the C4N atom of the nicotinamide ring suggest that Ca^{2+} disturbs the optimal orientation (distance and angle) of the reacting molecular orbitals for hydride transfer (*28*). Yet, despite the higher ionic radius of Ca^{2+} (0.99 Å) vs. that of Mg^{2+} (0.65 Å), our results show that Ca^{2+} in the "closed" [Ca^{2+}:ICT:$NADP^+$] pseudo-Michaelis complex is geometrically equivalent to the cocatalytic Mg^{2+} metal ion observed in the "closed" K100M [Mg^{2+}:ICT:$NADP^+$] x2 ternary product complex (Fig. 4.8A). Being heavier than Mg^{2+}, Ca^{2+} might instead slow activity by affecting protein dynamics, perhaps by hindering mobility at the metal binding site during hinge closure. Kinetic data from EcoIDH (*37*) and pig heart IDH (*13*) and on the related homoisiocitrate dehydrogenase in yeast (*11*) support a steady-state kinetic mechanism with catalysis at least 10 times faster than the structural changes needed to set up the active site for catalysis.

The [Ca^{2+}:ICT:$NADP^+$] *pseudo*-Michaelis complex and the K100M [Mg^{2+}:ICT:$NADP^+$] x2 ternary product complex portray structural changes in the active site during the conversion of isocitrate to α-ketoglutarate (Fig.4.8A). The most significant conformational changes take place at Tyr160 and Glu336. While the change at Tyr160 arises following β-decarboxylation, that at Glu336 likely results from substituting the amine group of Lys100 by a water molecule or a NH_4^+ cation (W66).

Catalytic mechanism

Our structural data support, in detail, the catalytic mechanism proposed by Aktas and Cook (*6*). Kinetic studies of the related homoisocitrate dehydrogenase (HIDH) (*61*) suggest that an invariant Lys (Lys230* in *Eco*IDH) initiates dehydrogenation by abstracting the proton from the hydroxyl of isocitrate. Consistent with this proposal, and seen for the first time, Lys230* forms a hydrogen bond with the substrate hydroyxl in the fully closed [Ca^{2+}:ICT:$NADP^{+}$] *pseudo*-Michaelis complex. Lys230* must be uncharged in order to act as a base. As predicted (*6*) Asp307, a metal ligand, aids proton abstraction from the hydroxyl by hydrogen bonding to Lys230*. Conversion of isocitrate into oxalosuccinate is fast (*37*) and, except for the observed ~90º rotation of the C1-carboxylate upon the change in hybridization at C2 from sp^3 to sp^2, need be accompanied only by minimal structural rearrangements in the active site pocket.

Next, decarboxylation of oxalosuccinate produces an enol intermediate, with the C2 carbonyl re-protonated by Lys230*. Loss of CO_2 allows an invariant Tyr (Tyr160 in *Eco*IDH) to approach C3 for stereospecific protonation, while Lys230* once again abstracts the proton from the C2 hydroxyl to yield the α-ketoglutarate product. In the K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 ternary product complex, Lys230* hydrogen bonds to the C2 carbonyl while Tyr160 seems poised to (de)protonate C3, albeit 3.4 Å away. Indeed, the near planar α-ketoglutarate suggests some double-bond character between C2 and C3 (reducing the C1-C2-C3-C4 dihedral angle from 20° to 0° does not greatly compromise the observed electron density). Alternatively, the electron density may represent α-hydroxyglutarate (and $NADP^{+}$). In the absence of CO_2, IDH uses NADPH to reduce α-ketoglutarate to α-hydroxyglutarate, and NMR studies have detected the signature of α-hydroxyglutarate bound to pig heart IDH (*12*). Thus, the observed electron density is probably an average dominated by the carbonyl form of α-ketoglutarate with lesser contributions from the enol form and α-hydroxyglutarate.

Aktas and Cook (*6*) proposed that the IDH mechanism is similar to that of the malic enzyme, i.e., malate dehydrogenase (decarboxylating). In the malic enzyme a hydrogen bond exists between the catalytic Lys and Tyr, allowing dehydrogenation and decarboxylation to proceed without the need to exchange protons with other groups or with bulk solvent. Neither in the [Ca^{2+}:ICT:$NADP^{+}$] *pseudo*-Michaelis complex nor in the K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 ternary product complex is there evidence of a hydrogen bond between Lys230* and Tyr160, even though their functional groups lie in close proximity. However, both hydrogen bond to Asp307 in the [Ca^{2+}:ICT:$NADP^{+}$] *pseudo*-Michaelis complex. If a proton is not exchanged directly between Lys230* and Tyr160 during enol conversion to the α-ketoglutarate, then the exchange may be indirect via Asp307. Alternatively, extending from the Lys230*-Asp307-Tyr160 catalytic triad there is a proton relay comprising two water molecules (W5 and W4), the O2N oxygen from the NADP pyrophosphate backbone and a third water molecule (W6) that may balance proton flow between the active site and the bulk solvent during catalysis (Fig. 4.10A and Fig. 4.10B). Although a proton relay of this kind has been proposed for the *pseudo*-Michaelis complex of *Ape*IDH (*21*), our results provide new insights into the role played by the Lys230*-Asp307-Tyr160 catalytic triad during oxidative decarboxylation.

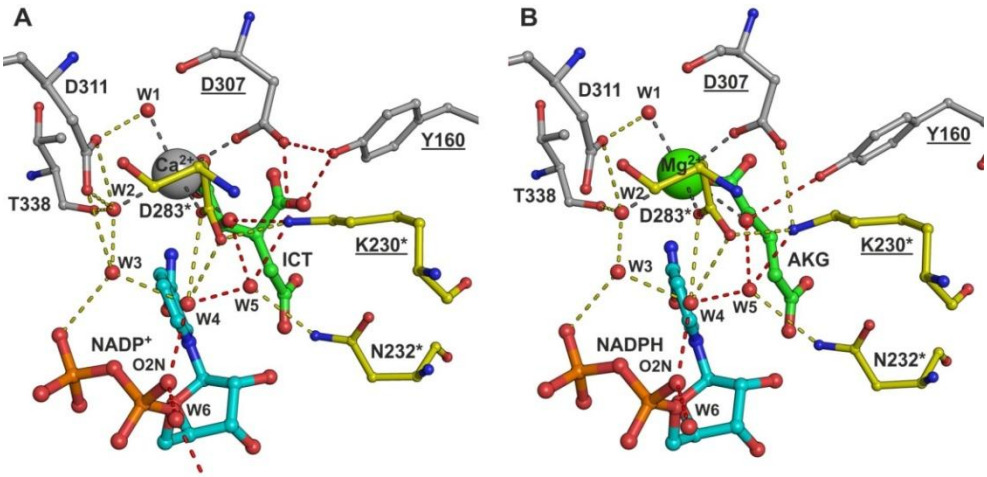


Figure 4.10 The catalytic triad and the proton relay. Representation of the structural and electrostatic readjustments that take place at the catalytic site during the conversion of ICT into α-KG. (**A**) The *pseudo*-Michaelis complex of the [Ca^{2+}:ICT:$NADP^{+}$] structure mimics the dehydrogenation stage of the C2-hydroxyl group of ICT, mediated by Lys230*. (**B**) The ternary complex product trapped in the K100M [Mg^{2+}:ICT:$NADP^{+}$] x2 structure, mirrors the hydrogenation stage at the C2 carbon of OXA mediated by Tyr160. Dashed lines indicate hydrogen or coordination bonds between protein residues, water molecules and ligands, and are color-coded according to their role: red for the proton relay involved in the acid/base assisting catalysis, yellow for the interactions between protein residues and the conserved water molecules, and gray for the interactions involved in metal coordination. The protein residues, the $NADP^{+}$/NADPH and the ICT/AKG ligands are represented in ball-and-stick, and the water molecules and the Ca^{2+} ion are shown as spheres. Atom colors are: gray for calcium; orange for phosphorus, blue for nitrogen and red for oxygen; carbon atoms are colored cyan for $NADP^{+}$/NADPH, green for ICT/AKG, and light gray and yellow for protein carbon atoms of the first and second subunits of *Eco*IDH, respectively. The labels of the protein residues from the second subunit are suffixed with an asterisk and the residues Lys230*-Tyr160-Asp307 in the catalytic triad are highlighted with underlined labels.

Table 4.1. Crystallization and crystal soaking procedures for all datasets, and their respective bound ligands [a]

Dataset Name	Resolution (Å)	Solution 1	Solution 2	Soak Time (min)	Final pH	Reaction products	Bound ligands
[Ca^{2+}:ICT:$NADP^{+}$]	1.9	300 *m*M ICT, 52 *m*M Ca^{2+}	300 *m*M ICT, 500 *m*M $NADP^{+}$, 52 *m*M Ca^{2+} [b]	60	7.0	-	Ca^{2+}:ICT , $NADP^{+}$
K100M [Mg^{2+}:ICT:$NADP^{+}$] x1	1.8	300 *m*M ICT, 52 *m*M Mg^{2+}	300 *m*M ICT, 500 *m*M $NADP^{+}$, 52 *m*M Mg^{2+} [b]	90	7.5	-	Mg^{2+}:ICT, NMN, A2P
K100M [Mg^{2+}:ICT:$NADP^{+}$] x2	2.7	N/A [c]	300 *m*M ICT, 500 *m*M $NADP^{+}$, 52 *m*M Mg^{2+} [b, c]	150	7.5	α-KG, NADPH	Mg^{2+}: α-KG, NADPH, NMN
IDHK100M [Ca^{2+}:α-KG:NADPH]	2.3	N/A	77 *m*M ICT, 10 *m*M NADPH, 8.5 *m*M Ca^{2+} [c, d]	30	8.0	-	Ca^{2+}: α-KG, A2P
[Ca^{2+}:ICT:thio-$NADP^{+}$]	1.8	300 *m*M ICT, 52 *m*M Ca^{2+}	300 *m*M ICT, 500 *m*M $NADP^{+}$, 52 *m*M Ca^{2+} [b]	50	7.5	-	Ca^{2+}: ICT; thio-$NADP^{+}$
K100M [Mg^{2+}:ICT:thio-$NADP^{+}$]	1.8	300 *m*M ICT, 52 *m*M Mg^{2+}	300 *m*M ICT, 400 *m*M $NADP^{+}$, 52 *m*M Mg^{2+} [b]	50	7.5	-	Mg^{2+}: ICT; thio-$NADP^{+}$

[a] Only the substrates and/or substrate analogues added to the soaking solutions are described. The soaks comprised either one (Solution 2) or two steps (Solutions 1 and 2) depending on the dataset; [b] the dilution factor (D.F. ≅ 1.6) resulting from the dissolution of the cofactor and the extra addition of the NH_3OH base at 20% is not considered. [c] Crystal from a "dry drop" was rehydrated with 1.87 M NH_4SO_4, 200 *m*M NaHEPES pH 7.5. [d] crystals dissolved in 5 minutes when replacing Ca^{2+} by Mg^{2+}.

Table 4.2. Data collection and processing statistics [a]

Data Sets	$[Ca^{2+}$:ICT:$NADP^+]$	K100M $[Mg^{2+}$:ICT:$NADP^+]$ x1	K100M $[Mg^{2+}$:ICT:$NADP^+]$ x2	K100M $[Ca^{2+}$:α-KG:NADPH]	K100M $[Mg^{2+}$:ICT:thio-$NADP^+]$	$[Ca^{2+}$:ICT:thio-$NADP^+]$
Beamline	In house	ESRF ID23-1	ESRF ID14-4	In house	SLS PXIII	SLS PXIII
Detector	Proteum Pt135	ADSC Q315r	ADSC Q315r	Proteum Pt135	Mar225	Mar225
Processing packages	ProteumSuite, XPREP, CCP4	XDS, CCP4	XDS, CCP4	Proteum Suite, XPREP, CCP4	XDS, CCP4	XDS, CCP4
Wavelength (Å)	1.5418	0.9795	0.9535	1.5418	0.9792	0.9792
Space group			*P* 4_32_12			
Cell parameters (Å)	a = b = 105.30 c = 145.72	a = b = 103.58 c = 149.68	a = b = 105.79 c = 145.49	a = b = 103.68 c = 149.58	a = b = 103.05 c = 150.34	a = b = 104.01 c = 149.56
Resolution (Å)	47.10-1.90 (2.00- 1.90)	50.00-1.80 (1.91-1.80)	60.00-2.69 (2.85-2.69)	46.37-2.30 (2.40-2.30)	48.74-1.90 (2.01-1.90)	50.0-1.80 (1.91-1.80)
Nr. observations	370445 (30740)	851107 (128548)	203269 (30161)	557631 (40200)	468880 (71214)	552349 (85244)
Nr. unique reflections	65181 (9149)	75372 (11833)	23556 (3636)	36981 (4332)	64372 (9910)	76263 (11806)
Completeness (%)	99.9 (100.0)	99.7 (98.5)	99.5 (97.5)	100.0 (100.0)	99.4 (96.1)	99.4 (96.7)
Multiplicity	5.7 (3.4)	11.3 (10.8)	8.6 (8.3)	15.1 (9.3)	7.3 (7.2)	7.2 (7.2)
<I/σ(I)>	11.2 (1.7)	12.3 (2.6)	15.2 (2.8)	24.8 (6.1)	18.2 (3.2)	17.7 (2.7)
R-merge (%) [b]	11.0 (54.0)	12.6 (70.8)	12.5 (83.5)	8.3 (31.4)	7.6 (54.5)	6.7 (62.6)
R-meas (%) [b]	-	13.1 (74.2)	13.3 (89.0)	-	8.2 (58.7)	7.2 (67.4)
R-pim (%) [b]	4.9 (34.8)	-	-	2.1 (10.8)	-	-
Wilson B ($Å^2$)	18.7	30.6	51.6	32.7	31.2	31.8

[a] Values in parentheses refer to the highest resolution shell; [b] R-merge = merging R-factor, $(\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|) / (\sum_{hkl} \sum_i I(hkl)) \times 100\%$; R-meas = multiplicity independent R-factor, $\sum_{hkl} [N_{hkl}/(N_{hkl} -1)]^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i (hkl) \times 100\%$ (*62*); R-pim = precision independent R-factor, $\sum_{hkl} [1/(N_{hkl} -1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i (hkl) \times 100\%$ (*62*), where $I_i(hkl)$ is the observed intensity, $\langle I(hkl)\rangle$ is the average intensity of multiple observations from symmetry-related reflections, and N_{hkl} is their multiplicity.

Table 4.3. Refinement Statistics

	[Ca^{2+}:ICT:NADP]	K100M [Mg^{2+}:ICT:$NADP^+$] x1	K100M [Mg^{2+}:ICT:$NADP^+$] x2
Resolution limits (Å)	44.81-1.90 (1.97-1.90)	46.34-1.80 (1.86-1.80)	52.90-2.68 (2.80-2.70)
R-factor [a]	0.190 (0.351)	0.170 (0.253)	0.160 (0.230)
nr. working reflections [b]	61806 (6069)	71597 (6867)	22359 (2362)
Free R-factor [c, d]	0.224 (0.347)	0.19 (0.284)	0.210 (0.292)
nr. free reflections [b]	3272 (324)	3768 (358)	1195 (119)
Overall coordinate error estimate (Å) [e]	0.58	0.38	0.66
Model completeness and composition (%)			
Regions omitted	-	Met 1	Met 1
Mg^{2+}/ Ca^{2+}	Ca^{2+}	Mg^{2+}	Mg^{2+}
$NADP^+$/NADPH /thio-$NADP^+$	$NADP^+$	-	NADPH
NMN moiety	-	1	1
A2P moiety	-	1	-
ICT/α-KG	ICT	ICT	α-KG
SO_4^{2-}	-	1	-
Loop conformation	"closed"	"quasi-closed"	"closed"
Non-hydrogen protein atoms	3674	3814	3412
Disordered residues	5	6	-
Solvent molecules	374	504	127
Mean B values (Å^2) [f]			
Protein main-chain	27.2	29.5	25.3
Protein side-chain	30.0	33.4	29.5
Ions and ligands	27.7	40.7	40.2
Solvent	38.7	40.1	27.4
Model r.m.s. deviations from ideality			
Bond lengths (Å)	0.007	0.007	0.008
Bond angles (°)	1.088	1.041	1.139
Chiral centers (Å^3)	0.070	0.072	0.069
Planar groups (Å)	0.004	0.004	0.004
Model validation [g]			
% Ramachandran outliers	0.0	0.0	0.5
% Ramachandran favored	96.9	97.6	96.1
% Rotamer outliers	0.0	0.9	2.4
C^β outliers	0	0	0
Clash score	6.00	7.01	9.37
PDB accession ID	4aj3	4ajs	4ajr

a R-factor = ∑hkl ||Fo|-|Fc|| / ∑hkl |Fo|, where |Fo| and |Fc| are the observed and calculated structure factor amplitudes, respectively; ***b*** No σ(Fo) cutoff ; ***c*** Cross-validation R-factor computed from a randomly chosen subset of 5% of the total number of reflections were not used during the refinement.; ***d*** *R*-free flag derived from [Ca^{2+}:ICT:$NADP^+$] structure ; ***e*** Maximum-likelihood estimate; ***f*** Model compositions is referred to each monomer***;*** ***f*** Calculated from isotropic or equivalent isotropic B-values; ***g*** Calculated with MolProbity (*48*).

Table 4.3. Refinement Statistics (*continuation*)

	K100M [Ca^{2+}:α-KG:NADPH]	K100M [Mg^{2+}:ICT:thio-NADP$^+$]	[Ca^{2+}:ICT:thio-NADP$^+$]
Resolution limits (Å)	44.30-2.30 (2.38-2.30)	48.90- 1.90 (1.97-1.90)	46.5-1.80 (1.86-1.80)
R-factor [a]	0.175 (0.233)	0.165 (0.242)	0.183 (0.270)
Nr. working reflections [b]	34197 (3324)	61051(5951)	72371 (6824)
Free R-factor [c, d]	0.210 (0.278)	0.181 (0.263)	0.204 (0.286)
nr. free reflections [b]	1822 (203)	3235 (308)	3802 (331)
Overall coordinate error estimate (Å) [e]	0.64	0.42	0.37
Model completeness and composition (%)			
Regions omitted	Met 1	Met 1	Met 1
Mg^{2+}/ Ca^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}
NADP$^+$/NADPH /thio-NADP$^+$	-	thio-NADP	thio-NADP
NMN moiety	-	-	-
A2P moiety	1	-	-
ICT/α-KG	α-KG	ICT	ICT
SO_4^{2-}	1	1	1
Loop conformation	"quasi-closed"	"quasi-closed"	"quasi-closed"
Non-hydrogen protein atoms	3716	3809	3818
Disordered residues	7	9	7
Solvent molecules	429	490	507
Mean B values (Å^2) [f]			
Protein main-chain	27.08	27.8	29.07
Protein side-chain	30.40	31.7	32.57
Ions and ligands	47.8	38.3	33.0
Solvent	42.9	42.4	42.7
Model r.m.s. deviations from ideality			
Bond lengths (Å)	0.008	0.007	0.007
Bond angles (º)	1.101	1.079	1.086
Chiral centers (Å^3)	0.073	0.073	0.075
Planar groups (Å)	0.005	0.005	0.004
Model validation [g]			
% Ramachandran outliers	0.0	0.0	0.0
% Ramachandran favored	96.9	96.9	97.1
% Rotamer outliers	0.9	0.3	0.3
C^β outliers	0	0	0
Clash score	8.81	5.55	6.30
PDB accession ID	4ajc	4ajb	4aja

a R-factor = ∑hkl ||Fo|-|Fc|| / ∑hkl |Fo|, where |Fo| and |Fc| are the observed and calculated structure factor amplitudes, respectively; ***b*** No σ(Fo) cutoff ; ***c*** Cross-validation R-factor computed from a randomly chosen subset of 5% of the total number of reflections were not used during the refinement.; ***d*** *R*-free flag derived from [Ca^{2+}:ICT:NADP$^+$] structure ; ***e*** Maximum-likelihood estimate; ***f*** Model compositions is referred to each monomer***;*** ***f*** Calculated from isotropic or equivalent isotropic B-values; ***g*** Calculated with MolProbity (*48*).

4.6 REFERENCES

1. Hurley, J. H., Thorsness, P. E., Ramalingam, V., Helmers, N. H., Koshland, D. E., Jr., and Stroud, R. M. (1989) Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase, *Proc Natl Acad Sci U S A 86*, 8635-8639.
2. Dean, A. M., and Golding, G. B. (1997) Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase, *Proc Natl Acad Sci U S A 94*, 3104-3109.
3. Garrad, R. C., and Bhattacharjee, J. K. (1992) Lysine biosynthesis in selected pathogenic fungi: characterization of lysine auxotrophs and the cloned LYS1 gene of *Candida albicans*, *J Bacteriol 174*, 7379-7384.
4. Tipton, P. A., and Beecher, B. S. (1994) Tartrate dehydrogenase, a new member of the family of metal-dependent decarboxylating R-hydroxyacid dehydrogenases, *Arch Biochem Biophys 313*, 15-21.
5. Salomone, J. Y., Crouzet, P., De Ruffray, P., and Otten, L. (1996) Characterization and distribution of tartrate utilization genes in the grapevine pathogen *Agrobacterium vitis*, *Mol Plant-Microbe Interact 9*, 401-408.
6. Aktas, D. F., and Cook, P. F. (2009) A lysine-tyrosine pair carries out acid-base chemistry in the metal ion-dependent pyridine dinucleotide-linked beta-hydroxyacid oxidative decarboxylases, *Biochemistry 48*, 3565-3577.
7. Steinberger, R., and Westheimer, F. H. (1951) Metal Ion-catalyzed Decarboxylation: A Model for an Enzyme System1, *J Am Chem Soc 73*, 429-435.
8. Lienhard, G. E., and Rose, I. A. (1964) The Mechanism of Action of 6-Phosphogluconate Dehydrogenase*, *Biochemistry 3*, 190-195.
9. Dalziel, K., and Londesborough, J. C. (1968) The mechanisms of reductive carboxylation reactions. Carbon dioxide or bicarbonate as substrate of nicotinamide-adenine dinucleotide phosphate-linked isocitrate dehydrogenase and malic enzyme, *Biochem J 110*, 223-230.

10. Siebert, G., Carsiotis, M., and Plaut, G. W. (1957) The enzymatic properties of isocitric dehydrogenase, *J Biol Chem 226*, 977-991.
11. Cook, P. F., and Cleland, W. W. (1981) pH variation of isotope effects in enzyme-catalyzed reactions. 2. Isotope-dependent step not pH dependent. Kinetic mechanism of alcohol dehydrogenase, *Biochemistry 20*, 1805-1816.
12. Ehrlich, R. S., and Colman, R. F. (1987) Ionization of isocitrate bound to pig heart $NADP^+$-dependent isocitrate dehydrogenase: ^{13}C NMR study of substrate binding, *Biochemistry 26*, 3461-3466.
13. Grissom, C. B., and Cleland, W. W. (1988) Isotope effect studies of the chemical mechanism of pig heart NADP isocitrate dehydrogenase, *Biochemistry 27*, 2934-2943.
14. Hurley, J. H., Dean, A. M., Koshland, D. E., Jr., and Stroud, R. M. (1991) Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and $NADP^+$ complexes, *Biochemistry 30*, 8671-8678.
15. Stoddard, B. L., Dean, A., and Bash, P. A. (1996) Combining Laue diffraction and molecular dynamics to study enzyme intermediates, *Nat Struct Biol 3*, 590-595.
16. Stoddard, B. L., Dean, A., and Koshland, D. E., Jr. (1993) Structure of isocitrate dehydrogenase with isocitrate, nicotinamide adenine dinucleotide phosphate, and calcium at 2.5-A resolution: a pseudo-Michaelis ternary complex, *Biochemistry 32*, 9310-9316.
17. Brubaker, M. J., Dyer, D. H., Stoddard, B., and Koshland, D. E., Jr. (1996) Synthesis, kinetics, and structural studies of a photolabile caged isocitrate: a catalytic trigger for isocitrate dehydrogenase, *Biochemistry 35*, 2854-2864.
18. Grodsky, N. B., Soundar, S., and Colman, R. F. (2000) Evaluation by site-directed mutagenesis of aspartic acid residues in the metal site of pig heart NADP-dependent isocitrate dehydrogenase, *Biochemistry 39*, 2193-2200.

19. Huang, Y. C., Grodsky, N. B., Kim, T.-K., and Colman, R. F. (2004) Ligands of the Mn^{2+} Bound to Porcine Mitochondrial NADP-Dependent Isocitrate Dehydrogenase, as Assessed by Mutagenesis, *Biochemistry 43*, 2821-2828.
20. Ceccarelli, C., Grodsky, N. B., Ariyaratne, N., Colman, R. F., and Bahnson, B. J. (2002) Crystal structure of porcine mitochondrial $NADP^+$-dependent isocitrate dehydrogenase complexed with Mn^{2+} and isocitrate. Insights into the enzyme mechanism, *J Biol Chem 277*, 43454-43462.
21. Karlstrom, M., Stokke, R., Steen, I. H., Birkeland, N. K., and Ladenstein, R. (2005) Isocitrate dehydrogenase from the hyperthermophile *Aeropyrum pernix*: X-ray structure analysis of a ternary enzyme-substrate complex and thermal stability, *J Mol Biol 345*, 559-577.
22. Lee, M. E., Dyer, D. H., Klein, O. D., Bolduc, J. M., Stoddard, B. L., and Koshland, D. E., Jr. (1995) Mutational analysis of the catalytic residues lysine 230 and tyrosine 160 in the $NADP^+$-dependent isocitrate dehydrogenase from *Escherichia coli*, *Biochemistry 34*, 378-384.
23. Hurley, J. H., Dean, A. M., Sohl, J. L., Koshland, D. E., Jr., and Stroud, R. M. (1990) Regulation of an enzyme by phosphorylation at the active site, *Science 249*, 1012-1016.
24. Bolduc, J. M., Dyer, D. H., Scott, W. G., Singer, P., Sweet, R. M., Koshland, D. E., Jr., and Stoddard, B. L. (1995) Mutagenesis and Laue structures of enzyme intermediates: isocitrate dehydrogenase, *Science 268*, 1312-1318.
25. Chen, R., Dean, A. M., Grobler, J. A., and Hurley, J. H. (1996) Second-site suppression of regulatory phosphorylation in *Escherichia col*i isocitrate dehydrogenase, *Protein Sci 5*, 287-295.
26. Cherbavaz, D. B., Lee, M. E., Stroud, R. M., and Koshland, D. E., Jr. (2000) Active site water molecules revealed in the 2.1 A resolution structure of a site-directed mutant of isocitrate dehydrogenase, *J Mol Biol 295*, 377-385.

27. Imada, K., Inagaki, K., Matsunami, H., Kawaguchi, H., Tanaka, H., Tanaka, N., and Namba, K. (1998) Structure of 3-isopropylmalate dehydrogenase in complex with 3-isopropylmalate at 2.0 A resolution: the role of Glu88 in the unique substrate-recognition mechanism, *Structure 6*, 971-982.
28. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences, *Science 277*, 202-206.
29. Stoddard, B. L., Cohen, B. E., Brubaker, M., Mesecar, A. D., and Koshland, D. E., Jr. (1998) Millisecond Laue structures of an enzyme-product complex using photocaged substrate analogs, *Nat Struct Biol 5*, 891-897.
30. Doyle, S. A., Beernink, P. T., and Koshland, D. E., Jr. (2001) Structural basis for a change in substrate specificity: crystal structure of S113E isocitrate dehydrogenase in a complex with isopropylmalate, Mg^{2+}, and NADP, *Biochemistry 40*, 4234-4241.
31. Peng, Y., Zhong, C., Huang, W., and Ding, J. (2008) Structural studies of *Saccharomyces cerevesiae* mitochondrial NADP-dependent isocitrate dehydrogenase in different enzymatic states reveal substantial conformational changes during the catalytic reaction, *Prot Sci 17*, 1542-1554.
32. Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liau, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate, *Nature 462*, 739-744.
33. Yang, B., Zhong, C., Peng, Y., Lai, Z., and Ding, J. (2010) Molecular mechanisms of "off-on switch" of activities of human IDH1 by tumor-associated mutation R132H, *Cell Res 20*, 1188-1200.

34. Malik, R., and Viola, R. E. (2010) Structural characterization of tartrate dehydrogenase: a versatile enzyme catalyzing multiple reactions, *Acta Cryst D 66*, 673-684.
35. Kim, T.-K., Lee, P., and Colman, R. F. (2003) Critical Role of Lys212 and Tyr140 in Porcine NADP-dependent Isocitrate Dehydrogenase, *J Biol Chem 278*, 49323-49331.
36. Miller, S. P., Lunzer, M., and Dean, A. M. (2006) Direct demonstration of an adaptive constraint, *Science 314*, 458-461.
37. Dean, A. M., and Koshland, D. E., Jr. (1993) Kinetic mechanism of *Escherichia coli* isocitrate dehydrogenase, *Biochemistry 32*, 9302-9309.
38. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem 72*, 248-254.
39. Kabsch, W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants, *J Appl Crystallogr 26*, 795-800.
40. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Cryst D 67*, 235-242.
41. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions, *Acta Cryst D 60*, 432-438.
42. Matthews, B. W. (1968) Solvent content of protein crystals, *J Mol Biol 33*, 491-497.
43. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Cryst D 66*, 213-221.

44. Vagin, A. A., Murshudov, G. N., and Strokopytov, B. V. (1998) BLANC: the program suite for protein crystallography, *J Appl Crystallogr 31*, 98-102.
45. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Cryst D 66*, 486-501.
46. Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A., and Murshudov, G. N. (2012) JLigand: a graphical tool for CCP4 template restraint library *Acta Cryst 68,*431-440.
47. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) ARP/wARP and molecular replacement, *Acta Cryst D 57*, 1445-1450.
48. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst D 66*, 12-21.
49. Velankar, S., Best, C., Beuth, B., Boutselakis, C. H., Cobley, N., Sousa Da Silva, A. W., Dimitropoulos, D., Golovin, A., Hirshberg, M., John, M., Krissinel, E. B., Newman, R., Oldfield, T., Pajon, A., Penkett, C. J., Pineda-Castillo, J., Sahni, G., Sen, S., Slowley, R., Suarez-Uruena, A., Swaminathan, J., van Ginkel, G., Vranken, W. F., Henrick, K., and Kleywegt, G. J. (2010) PDBe: Protein Data Bank in Europe, *Nucl Acids Res 38*, D308-D317.
50. Kabsch, W. (1976) Solution for Best Rotation to Relate 2 Sets of Vectors, *Acta Cryst A 32*, 922-923.
51. Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme, *Proteins 30*, 144-154.
52. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein alignment in three dimensions, *Acta Cryst D 60*, 2256-2268.

53. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta CrystD 59*, 1131-1137.
54. Finer-Moore, J., Tsutakawa, S. E., Cherbavaz, D. B., LaPorte, D. C., Koshland, D. E., and Stroud, R. M. (1997) Access to Phosphorylation in Isocitrate Dehydrogenase May Occur by Domain Shifting, *Biochemistry 36*, 13890-13896.
55. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proc Natl Acad Sci U S A 98*, 10037-10041.
56. Sidhu, N. S., Delbaere, L. T., and Sheldrick, G. M. (2011) Structure of a highly NADP(+)-specific isocitrate dehydrogenase, *Acta Cryst D 67*, 856-869.
57. Hurley, J. H., and Dean, A. M. (1994) Structure of 3-isopropylmalate dehydrogenase in complex with NAD^+: ligand-induced loop closing and mechanism for cofactor specificity, *Structure 2*, 1007-1016.
58. Xu, X., Zhao, J., Xu, Z., Peng, B., Huang, Q., Arnold, E., and Ding, J. (2004) Structures of Human Cytosolic NADP-dependent Isocitrate Dehydrogenase Reveal a Novel Self-regulatory Mechanism of Activity, *J Biol Chem 279*, 33946-33957.
59. Cozzone, A. J., and El-Mansi, M. (2005) Control of Isocitrate Dehydrogenase Catalytic Activity by Protein Phosphorylation in *Escherichia coli*, *J Mol Microbiol Biotechnol 9*, 132-146.
60. Zheng, J., and Jia, Z. (2010) Structure of the bifunctional isocitrate dehydrogenase kinase/phosphatase, *Nature 465*, 961-965.
61. Lin, Y., West, A. H., and Cook, P. F. (2009) Site-directed mutagenesis as a probe of the acid-base catalytic mechanism of homoisocitrate dehydrogenase from *Saccharomyces cerevisiae*, *Biochemistry 48*, 7305-7312.

62. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat Struct Biol 4*, 269-275.

4.7 ACKNOWLEDGMENTS

The authors thank the European Synchrotron Radiation Facility (Grenoble, France) for support with the ID14-4 and ID23-1 data collections, the PXIII beamline of the Swiss Light Source (SLS), Paul Scherrer Institute in Villigen, Switzerland. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n.° 226716, and Fundação para a Ciência e Tecnologia (Portugal) grants PEst-OE/EQB/LA0004/2011 and SFRH/BD/23222/2005 to SG.

CHAPTER 5

GENERAL DISCUSSION

With the advances of the biotechnological industry, the role of compatible solutes as potential use in health-care, cosmetic and food applications has been one the major current topics of investigation. Recently, extensive reviews covering the physiological aspects of accumulation of these compounds (*1*), as well as the enzymology behind their synthesis as potential routes for production of glycosidic solutes as industrial fine chemicals (*2*) have been published. Indeed, protein engineering is currently pursued in different laboratories with the aim of broadening the donor and acceptor substrate specificities of naturally existing enzyme forms, to eventually generate a toolbox of new catalysts for glycoside synthesis (*3*).

Chapters 2 and 3, provide principles on the molecular mechanisms behind the synthesis of α-mannosylglycerate (MG), a sugar-derivative compatible solute with potential biotechnological interest.

MpgS adopts a central position in the biosynthetic two-step pathway of MG by catalyzing regio- and stereoselective formation of the O-glycosidic linkage between glycosyl-moiety from GDP-Man and 3-PG acceptor. MpgP completes the repertoire by hydrolysing the phosphoryl-intermediate mannosyl-3-phosphoglycerate (Fig. 1.1). The activities of glycosyltransferase and phosphatase are generally assumed to be under transcriptional control. There is currently no evidence supporting allosteric control over their activities. The NDP-sugar donor substrates and the phosphorylated or unphosphorylated acceptor substrates are recruited from central metabolic shunts of the cell. Availability of donor or acceptor could therefore present a possible alternative mechanism of regulation of biosynthesis of the glycosidic solutes. However, in order to delineate such a mechanism, evidence from a global ‘‘metabolomic-type’’ analysis of the response of the cell to salt would be required (*2*).

Along with the elucidation of the enzymological aspects, the deciphering of the molecular events behind the physiological response to environmental stress is equally important for global understanding of the regulatory mechanisms of the catalytic machinery for solute accumulation.

The structure of MpgS from *T. thermophilus* HB27 has brought novel insights into the elucidation of the molecular mechanism of the retaining glycosyl-transfer catalysis. Together with its MpgS homologue from *P. horikoshii*, both comprise the two structural representatives of the GT55 family. The MpgS crystallizes with a typical catalytic scaffold that favors the unique binding-mode of GDP-Man, distinctive from that commonly observed in other GTs, including their closest counterparts; the glucosyl-3-phosphoglycerate synthases (GpgS; GT81 family) and mannosylglycerate synthase from *R. marinus* [(MgS; GT78 family) (Fig. 2.6)]. This unique feature suggests a possible role for Asp167 in the DXD motif as the catalytic nucleophile, in light of both the S_N2-like double displacement and the $D_N{*}A_{Nss}$ mechanisms (Fig. 1.4). Replacement of Asp100 (equivalent to Asp167 in MpgS), in MgS resulted in almost complete lack of enzymatic activity (*4*).Though, in the majority of the known GTs structures in complex with their NDP-sugar donor or analogue, this residue (or it´s equivalogue) stabilizes instead the O3` hydroxyl group from the sugar moiety (Fig. 2.6; Table 2.4). Such binding mode may be the preferred one, amongst a variety of alternative conformations, for most of the crystallized retaining GTs. Yet, a reorientation of the sugar-moiety is most likely to follow the structural rearrangements that should take place during the binding of the acceptor and positioned in proximity to the scissile glycosidic bond. Typically, the binding of metal ion and sugar donor to the GT-A fold is accompanied by local conformational changes of flexible loop L2 (MpgS nomenclature) as, exemplified by loop rearrangements observed in Fig 2.13, and followed by additional flexible loop movement (loop L1) upon the acceptor binding as previously reported (*5, 6*). When binding the donor and acceptor, a cascade of electrostatic interactions should trigger structural rearrangement at the C-terminal region that sculpts the acceptor molecule binding site.

Given the vast structural diversity of acceptors, it is not surprising that the sequences and structures of the acceptor binding regions are divergent. In the GT-A fold, the central β-sheet core of the C-terminus is topologically conserved, but significant variation in C-terminal β-strands (α/β) [GnT-I (*7*)], α-helices [LgtC (*8*) and Lgt1 (*5*)] and/or loops [β-1,4-Gal-T1(*6*) and β-1,4-Gal-T7(*9*)] contributes to acceptor specificity and regiospecificity. The MpgS, MgS and GpgS, all comprise an α-helical content at their C-terminal region. Substrate and metal promiscuity has been demonstrated from MgS (*4, 10*), and for the MpgS/GpgS from *Rubrobacter xylanophilus* which was proposed to be at the evolutionary "cusp" whereby all MpgSs and GpgSs have evolved (*11*).

Structural and biochemical evidence, demonstrated for the first time the presence of a second catalytic metal ion in a retaining GT. In the GpgSs and MgS, this site is replaced by an Arg and Lys, respectively, thus strengthening the notion for stereotypically placed positive charge with a role in catalysis.

Despite of the amount of information that was given by solving of *T. thermophilus*` MpgS structure and by its comparison with its closest relatives, the full elucidation of the molecular mechanism for glycosyl transfer is still to be uncovered. Efforts in this direction were made recently for the GT-B retaining α,α-threhalose synthase (OstA; family GT20) via crystallization of a non-hydrolysable ternary complex analogue (*12*). Though, the central question regarding to the true nature of the catalytic nucleophile is still open Major drawback is the lack of a Michaelis` complex bound to a catalytically competent active site.

The structural elucidation of *T. thermophilus*` MpgP has helped to enlighten the basis of the molecular mechanism of phosphoryl-transfer reaction. The crystallographic "trapping" of several conformational states in a time scale as the formation of several reaction coordinate species, was determinant in the successful elucidation of the catalytic cycle (Fig 3.16). A metaphosphate-like species "trapped" at the active site provided structural evidence for a dissociative D_NS_N mechanism with a slight nucleophilic

character (Fig. 3.12 and Fig. 1.5). Such mechanism comes in agreement with previous proposals based on non-enzymatic phospho-monoester hydrolysis studies (*13, 14*), a paradigm which can be applied to other members of the HAD-like superfamily, and particularly to members that fall into the Cof-type clade. Cof-type phosphatases contain a remarkable diversification of their C2 cap through accretion of secondary structure elements to the basic unit with a three-stranded anti-parallel β-sheet (*15*). This clearly illustrates the successful "fitness" towards to exploration of the reaction spaces, by the combination of well preserved "HAD-like core" catalytic machinery with specialized modules tailored for substrate specificity.

Interestingly, the unrooted phylogenetic tree based on the amino acid sequence alignment of functionally characterized MpgPs and GpgPs and homologues, predicts three ill-defined clusters: one containing the MpgPs of *T. thermophilus* and *P. horikoshii* along with several homologues from (hyper)thermophilic organisms; a second cluster comprising known GpgPs and homologues, and a third one containing both MpgPs and GpgPs (Fig. 5.1). This convoluted distribution suggests the existence of a common ancestor with the ability to recognize both mannosyl-3-phosphoglycerate and glucosyl-3-phosphoglycerate as substrates. The topology of this phylogenetic tree contrasts clearly with that of MpgS/GpgS in which the clusters for MpgSs and GpgSs are far apart, in line with their strict substrate specificity.

The third topic discussed in this work focuses on the molecular determinants of activity-control in the $NADP^+$-isocitrate dehydrogenase from *E.coli*. In the active state, *Eco*IDH adopts a full "closed" conformation, unique from other conformational states known so far for this enzyme. Such conformation is favored with the binding of the substrate and $NADP^+$ cofactor, and a fine alignment of the *re*-face of nicotinamide moiety with the transferring hydride from C2 carbon atom of isocitrate is determinant for the enzymatic proficiency. Many other authors` have addressed this question (*16-18*), though they all lacked definitive structural evidence for a Michaelis ternary complex on these structures. Herein, we provide for the first time, "snapshots" of the

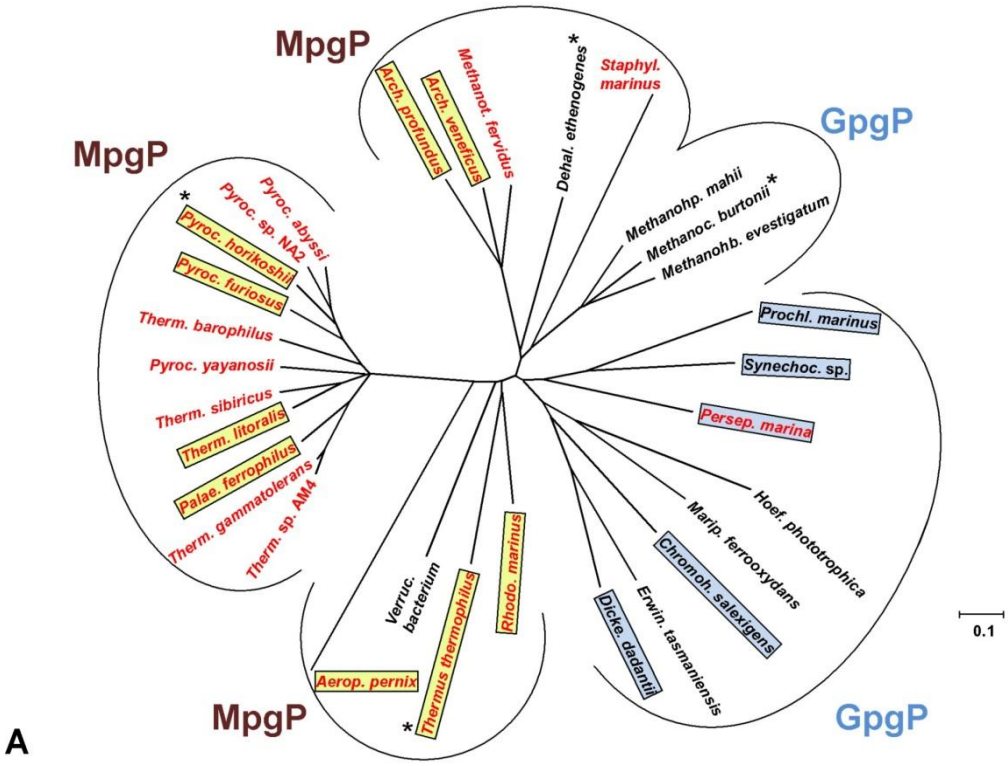
dehydrogenation and hydrogenations steps (Fig. 1.6) of *Eco*IDH β-oxidative decarboxylation reaction. Emphasis is given for the residue tetrad; Lys100-Leu103-Asn115-Glu336, and to the movement upwards the active centre by the "phosphorylation loop". The nature of these interactions plays pivotal role for the assemblage of the catalytic competent active-site, as the participating residues are strictly conserved amongst the functionally classified IDHs. Notably, in isopropylmalate dehydrogenase (IPMD) counterparts the correspondent tetrad is replaced by Leu-Val-Leu-Glu motif, demarking an evolutionary constrain towards selectivity to the substrate, isopropylmalate. Also, the presence of a conserved Glu in IPMDs, at a topological equivalent region as Ser113, i.e, at the N-terminal region of helix "d" in *Eco*IDH, may replace the now missing electrostatic interaction established between the γ-carboxylate group of isocitrate with the positive charge of the amine group of the nicotinamide pyridine ring. Whether Lys100 in *Eco*IDH has also chemical meaning as "electron driving force" during hydride transfer stage is still questionable. Furthermore, structural evidence for the diad Tyr160-Lys230* as assisting acid-base catalysts, is entirely in accord with their roles postulated Lin and Cook (*19*), together with a fine-tuned proton relay assembled in the catalytically competent active site (Fig. 4.10).

Three different examples of concerted protein dynamics coupled with catalytic activity were herein given for a glycosyltransferase, a phosphohydrolase and an oxidorreductase. The correlated motions of several loops and hinge rotations with the assemblage of the reactive substrate complexes were highlighted, as well as the importance of networks of coupled interactions that connect surface regions to active-site residues.

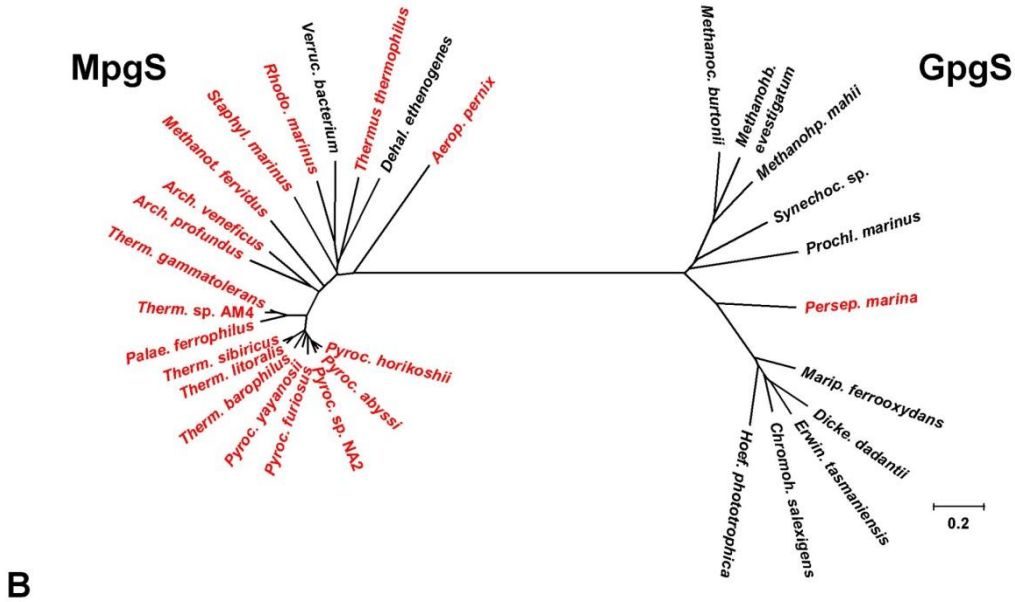
Factors others than protein flexibility with a role in enzyme catalysis are also discussed like; the electrostatic effect of solvent reorganization as in *Eco*IDH, the active-site architecture which is set up to preferentially bind the substrate in a conformation vicinal to its transition state (TS) as in MpgP, and to simultaneously allow a complementary environment for the stabilization of the TS.

Current debate on the relationship between protein dynamics and function, argues in favour as these being rooted in the free energy landscape of the enzyme, and that fluctuations at equilibrium can influence biological function or the evolution of a mechanism (*20-22*). This is highly suggestive in the case of the alternative mechanistic strategies proposed for the inverting and retaining glycosyltransferases.

Structural analysis in these three enzymatic systems supports the notion that selective advantage for a mechanism depends on the local energy minima provided by a strategic landscape. This will provide additional measurable data on the alternative enzymatic conformational states during the course of the reaction coordinate, and solidify the experimental background for its application into theoretical calculations on proteins` structural dynamics. Furthermore, the elucidation of the structural determinants behind the catalytic activity will provide guiding lines in the rational drawing of novel enzymes tailored for their industrial needs and with enhanced catalytic proficiency.


MpgP
MpgP
MpgP
GpgP
GpgP
Arch. profundus
Arch. veneficus
Methanot. fervidus
Dehal. ethenogenes *
Staphyl. marinus
Methanohp. mahii
Methanoc. burtonii *
Methanohb. evestigatum
Prochl. marinus
Synechoc. sp.
Persep. marina
Hoef. phototrophica
Marip. ferrooxydans
Chromoh. salexigens
Erwin. tasmaniensis
Dicke. dadantii
Rhodo. marinus
Verruc. bacterium
Thermus thermophilus
Aerop. pernix
Pyroc. abyssi
Pyroc. sp. NA2
Pyroc. horikoshii
Pyroc. furiosus
Therm. barophilus
Pyroc. yayanosii
Therm. sibiricus
Therm. litoralis
Palae. ferrophilus
Therm. gammatolerans
Therm. sp. AM4
0.1

A


MpgS
GpgS
Verruc. bacterium
Rhodo. marinus
Staphyl. marinus
Methanot. fervidus
Arch. veneficus
Arch. profundus
Therm. gammatolerans
Therm. sp. AM4
Palae. ferrophilus
Therm. sibiricus
Therm. litoralis
Therm. barophilus
Pyroc. yayanosii
Pyroc. furiosus
Pyroc. horikoshii
Pyroc. abyssi
Pyroc. sp. NA2
Thermus thermophilus
Dehal. ethenogenes
Aerop. pernix
Methanoc. burtonii
Methanohb. evestigatum
Methanohp. mahii
Synechoc. sp.
Prochl. marinus
Persep. marina
Marip. ferrooxydans
Dicke. dadantii
Erwin. tasmaniensis
Chromoh. salexigens
Hoef. phototrophica
0.2

B

Figure 5.1. Unrooted phylogenetic trees based on the amino acid sequences of (A) MpgPs, GpgPs and homologues, and (B) MpgSs, GpgSs and homologues. The classified bifunctional MpgS/GpgS from *Rubrobacter xylanophilus* is not included but it was shown to be at the evolutionary split from which all MpgSs and GpgSs have evolved (*23*). MEGA 4.1 software (24) was used for sequence alignment and to generate the phylogenetic tree using the neighbour-joining method. The scale bar represents the number of substitutions per site. Mesophiles (black) and (hyper)thermophiles (red). Asterisks indicate organisms harboring MpgP or GpgP with dual MpgP/GpgP activities. Organisms known to accumulate MG and GG are respectively highlighted by yellow and light blue boxes. MpgP, mannosyl-3-phosphoglycerate phosphatase; MpgS, mannosyl-3-phosphoglycerate synthase; GpgS, glucosyl-3-phosphoglycerate synthase; GpgP, glucosyl-3-phosphoglycerate phosphatase. Abbreviations: *Rhodo. marinus*, *Rhodothermus marinus* DSM 4252 (GenPept accession no. of MpgP and MpgS are YP_003289791.1, YP_003289792.1, respectively); *Thermus thermophilus* HB 27 (YP_004564.1, YP_004563.1); *Arch. profundus*, *Archaeoglobus profundus* DSM 5631 (YP_003400327.1, YP_003400326.1); *Arch. veneficus*, *Archaeoglobus veneficus* SNP6 (YP_004341720.1, YP_004342045.1); *Therm. litoralis*, *Thermococcus litoralis* DSM 5474 (AAY44816.1, AAY44815.1), *Therm.* sp. AM4, *Thermococcus* sp. AM4 (ZP_04878526.1, ZP_04878794.1); *Therm. sibiricus*, Thermococcus sibiricus MM 739 (YP_002993666.1, YP_002993667.1); *Therm. barophilus*, *Thermococcus barophilus* MP (YP_004070225.1, YP_004070206.1); *Therm. gammatolerans*, *Thermococcus gammatolerans* EJ3 (YP_002959182.1, YP_002959183.1); *Pyroc.* sp. NA2, *Pyrococcus* sp. NA2 (YP_004424374.1, YP_004424373.1); *Pyroc. horikoshii*, *Pyrococcus horikoshii* OT3 (NP_142849.1, NP_142850.1); *Pyroc. furiosus*, *Pyrococcus furiosus* DSM 3638 (NP_578319.2, NP_578320.1); *Pyroc. yayanosii*, *Pyrococcus yayanosii* CH1 (YP_004623656.1, YP_004623655.1); *Pyroc. abyssi*, *Pyrococcus abyssi* GE5 (NP_126909.1, NP_126908.1); *Palae. ferrophilus*, *Palaeococcus*

ferrophilus DSM 13482 (AAY44814.1, AAY44813.1); *Aerop. pernix*, *Aeropyrum pernix* K1(NP_147576.2, NP_147578.2); *Methanot. fervidus*, *Methanothermus fervidus* DSM 2088 (YP_004004107.1, YP_004004106.1); *Dehal. ethenogenes*, *Dehalococcoides ethenogenes* 195 (YP_182074.1); *Staphyl. marinus*, *Staphylothermus marinus* F1 (YP_001040515.1, YP_001040514.1); *Verruc. bacterium*, *Verrucomicrobiae bacterium* GD1235 (ZP_05058787.1, ZP_05055633.1); *Methanoc. burtonii*, *Methanococcoides burtonii* DSM 6242 (GenPept accession no. of GpgP and GpgS are YP_565451.1, YP_565452.1, respectively); *Persep. marina*, *Persephonella marina* EX-H1 (ABX75858.1, YP_002731348.1); *Prochl. marinus*, *Prochlorococcus marinus* CCMP1375 (NP_875121.1, NP_875119.1); *Hoef. phototrophica*, *Hoeflea phototrophica* DFL-43 (ZP_02165262.1, ZP_02165261.1); *Erwin. tasmaniensis*, *Erwinia tasmaniensis* Et1/99 (YP_001907933.1, YP_001907934.1); *Marip. ferrooxydans*, *Mariprofundus ferrooxydans* PV-1 (ZP_01453386.1, ZP_01453385.1); *Methanohb. evestigatum*, *Methanohalobium evestigatum* Z-7303 (YP_003725866.1, YP_003725907.1); *Methanohp mahii*, *Methanohalophilus mahii* DSM 5219 (YP_003542090.1, YP_003542479.1); *Dicke. dadantii*, *Dickeya dadantii* Ech 586 (YP_003333633.1, YP_003333634.1); *Chromoh. salexigens*, *Chromohalobacter salexigens* DSM 3043 (YP_574650.1, YP_574649.1); *Synechoc. sp.*, *Synechococcus sp.* PCC7002 (YP_001735265.1, YP_001735263.1). Figure was made by Dr. Nuno Borges and was obtained from previous work published by Gonçalves *et al*, 2011 (*25*).

REFERENCES

1. Santos, H., Lamosa, P., Borges, N., Gonçalves, L. G., Pais, T. M., and M. V. Rodrigues, M. V. (2011) Organic Compatible Solutes of Prokaryotes that Thrive in Hot Environments: The Importance of Ionic Compounds for Thermostabilization, In *Extremophiles Handbook* (Horikoshi, K., Antranikian, G., Bull, A. T., Robb, F. T., and Stetter, K. O., Eds.), pp 497-520, Springer, Tokyo.
2. Luley-Goedl, C., and Nidetzky, B. (2011) Glycosides as compatible solutes: biosynthesis and applications, *Nat Prod Rep 28*, 875-896.
3. Luley-Goedl, C., and Nidetzky, B. (2010) Carbohydrate synthesis by disaccharide phosphorylases: Reactions, catalytic mechanisms and application in the glycosciences, *Biotechnology Journal 5*, 1324-1338.
4. Flint, J., Taylor, E., Yang, M., Bolam, D., Tailforf, L., Martinez-Fleites, C., Dodson, E., BG Davies, Gilbert, H., and Davies, G. (2005) Structural Dissection and high-throughput screening of mannosyglyecrate synthase, *Nat Struct Mol Biol, 12*, 608-614.
5. Lu, W., Du, J., Stahl, M., Tzivelekidis, T., Belyi, Y., Gerhardt, S., Aktories, K., and Einsle, O. (2010) Structural basis of the action of glucosyltransferase Lgt1 from Legionella pneumophila, *J Mol Biol 396*, 321-331.
6. Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. (2008) Structure and function of beta -1,4-galactosyltransferase, *Curr Drug Targets 9*, 292-309.
7. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J. M. (2000) X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily, *EMBO J 19*, 5269-5280.
8. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs, *Nat Struct Biol 8*, 166-175.

9. Ramakrishnan, B., and Qasba, P. K. (2010) Crystal structure of the catalytic domain of Drosophila beta1,4-Galactosyltransferase-7, *J Biol Chem 285*, 15619-15626.
10. Nielsen, M. M., Suits, M. D., Yang, M., Barry, C. S., Martinez-Fleites, C., Tailford, L. E., Flint, J. E., Dumon, C., Davis, B. G., Gilbert, H. J., and Davies, G. J. (2011) Substrate and metal ion promiscuity in mannosylglycerate synthase, *J Biol Chem 286*, 15155-15164.
11. Empadinhas, N., and da Costa, M. S. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea, *Environ Microbiol 13*, 2056-2077.
12. Errey, J. C., Lee, S. S., Gibson, R. P., Martinez Fleites, C., Barry, C. S., Jung, P. M., O'Sullivan, A. C., Davis, B. G., and Davies, G. J. (2010) Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors, *Angew Chem Int Ed Engl 49*, 1234-1237.
13. Cleland, W. W. (1990) Secondary 18O isotope effects as a tool for studying reactions of phosphate mono-, di-, and triesters, *FASEB J 4*, 2899-2905.
14. Cleland, W. W., and Hengge, A. C. (1995) Mechanisms of phosphoryl and acyl transfer, *FASEB J 9*, 1585-1594.
15. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, *J. Mol. Biol. 361*, 1003-1034.
16. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences, *Science 277*, 202-206.
17. Doyle, S. A., Beernink, P. T., and Koshland, D. E., Jr. (2001) Structural basis for a change in substrate specificity: crystal structure of S113E isocitrate dehydrogenase in a complex with isopropylmalate, Mg^{2+}, and NADP, *Biochemistry 40*, 4234-4241.

18. Bolduc, J. M., Dyer, D. H., Scott, W. G., Singer, P., Sweet, R. M., Koshland, D. E., Jr., and Stoddard, B. L. (1995) Mutagenesis and Laue structures of enzyme intermediates: isocitrate dehydrogenase, *Science 268*, 1312-1318.
19. Lin, Y., West, A. H., and Cook, P. F. (2009) Site-directed mutagenesis as a probe of the acid-base catalytic mechanism of homoisocitrate dehydrogenase from *Saccharomyces cerevisiae*, *Biochemistry 48*, 7305-7312.
20. Papaleo, E., Tiberti, M., Invernizzi, G., Pasi, M., and Ranzani, V. (2011) Molecular determinants of enzyme cold adaptation: comparative structural and computational studies of cold- and warm-adapted enzymes, *Curr Protein Pept Sci 12*, 657-683.
21. Villali, J., and Kern, D. (2010) Choreographing an enzyme's dance, *Curr Opin Chem Biol 14*, 636-643.
22. Nashine, V. C., Hammes-Schiffer, S., and Benkovic, S. J. (2010) Coupled motions in enzyme catalysis, *Curr Opin Chem Biol 14*, 644-651.
23. Empadinhas, N., Pereira,P. J., Albuquerque, L., Costa, J., Sa--Moura, B., Marques, A.T., Macedo-Ribeiro, S., da Costa, M.S. (2011) Functional and structural characterization of a novel mannosyl-3-phosphoglycerate synthase from Rubrobacter xylanophilus reveals its dual substrate specificity, *Mol Microbiol* 79, 76-93.
24. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol 24*, 1596-1599.
25. Goncalves, S., Esteves, A. M., Santos, H., Borges, N. and Matias, P.M. (2011) Three-dimensional structure of mannosyl-3-phosphoglycerate phosphatase from *Thermus thermophilus* HB27: a new member of the haloalcanoic acid dehalogenase superfamily, *Biochemistry* 50, 9551-9567.

****

Este trabalho teve o apoio financeiro da FCT e do POPH/FSE no âmbito do Quadro Comunitário de Apoio, Bolsa Nº SFRH/BD/23222/2005.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt