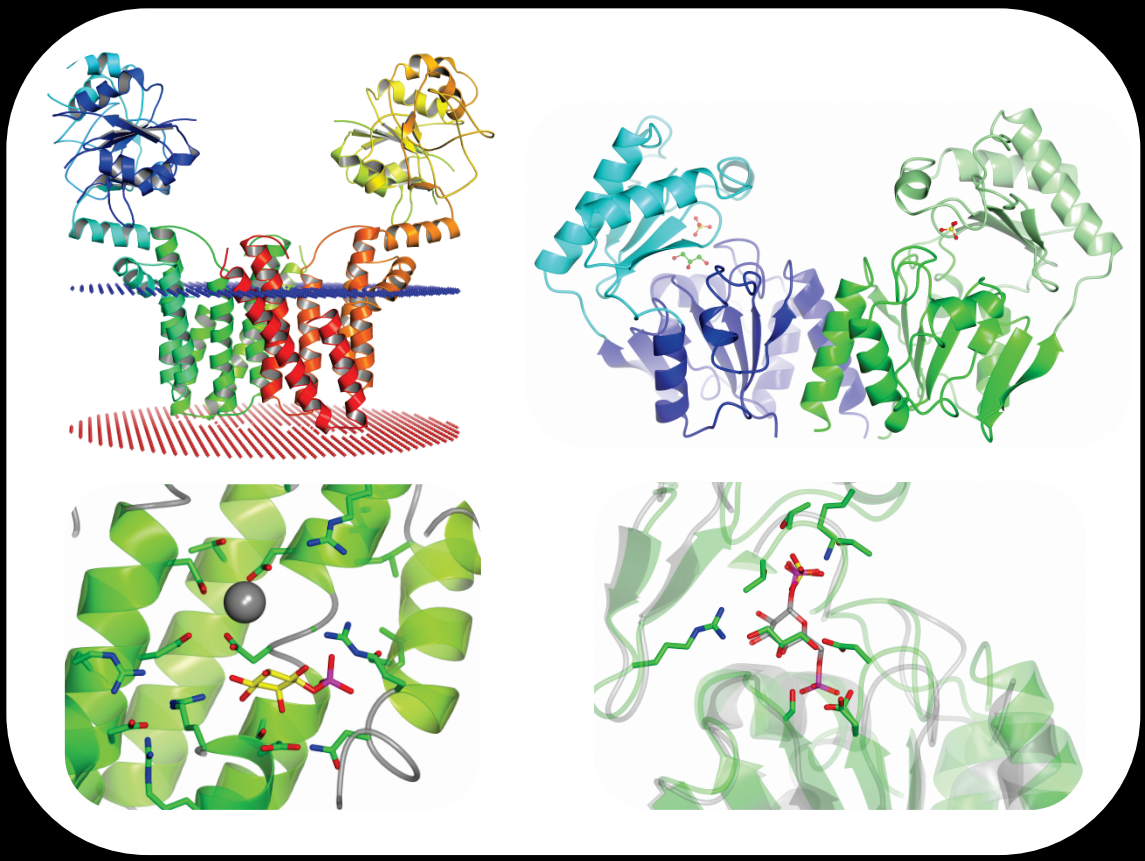
Crystallographic studies on membrane and cytoplasmic enzymes

Bifunctional cytidylyltransferase/CDP-alcohol phosphatidyltransferase and α-phosphoglucomutase

Przemyslaw Nogly



Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
April, 2013

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Crystallographic studies on membrane and cytoplasmic enzymes

Bifunctional cytidylyltransferase/CDP-alcohol phosphatidyltransferase and α-phosphoglucomutase

Przemyslaw Nogly

Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, April, 2013

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

ACKNOWLEDGMENTS

I would like to acknowledge my supervisor Dr Margarida Archer for giving me the exciting opportunity to work in structural biology shifting from my background in chemistry. Thanks for the friendly reception in her group, support in the research projects and in the final stages of this thesis writing, and sharing the insights into the structure analysis, scientific writing and science. I am thankful to be given the possibility to travel much during my PhD study seeking the knowledge needed to carry out the projects.

I owe my gratitude to many members of the Macromolecular Crystallography Unit who through eager discussions taught me the fundamentals and tricks of laboratory work and crystallography, which enabled concluding of this work.

I was very fortunate to have the company of Dr Miguel Pessanha to share the progress and misfortunes in studying difficult membrane proteins and with his supportive friendship I was able to carry on through the challenges of this PhD study. I am grateful for the generous sharing of the research experience and plentiful tips on the laboratory work received from Dr Matteo de Rosa, Dr Colin McVey, Dr Tiago Bandeiras, Dr Célia Romão, Catarina Silva and others, who have driven me to pay attention to experimental details and careful analysis of the results. I am indebted to Dr Pedro Matias, Dr Carlos Frazão and Dr Daniele de Sanctis for providing me insights into the crystallography and the valuable explanations I was given.

I am especially thankful to Dr Pikyee Ma for useful discussions and particularly the encouragement she has given to me in the final phase of my PhD study and during writing of this thesis.

I arrived to this stage only because of my parents Alicja and Krystian who all my life believed in me and encouraged me to always aim higher, despite difficult times and sacrifices. Therefore, I dedicate this thesis to them with all my love.

Finally I would like to acknowledge funding from the Marie Curie Initial Training Network ´Structural Biology of Membrane Proteins´ within the [European´s community´s] Seventh Framework Program [FP7/2007-2013] under grant agreement no [211800] and Fundação para a Ciência e a Tecnologia (FCT-69084, PTDC/BIA-MIC/099963/2008, PTDC/BIA-PRO/103718/2008, PTDC/SAU-NEU/103720/2008).

ABSTRACT

Living organisms perform chemical reactions in order to survive and to maintain functions from reproduction to cell growth. Their exposure to continuous changes in the environment and competitive pressure impose a requirement for rapid adjustment. These adaptive responses may include reduction of growth upon deprivation of a sugar source or minimization of damaging effects upon thermal stress. This PhD thesis includes the structural characterization of two enzymes, which respond to environmental changes.

The first protein studied within the scope of the PhD work is the bifunctional enzyme CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) and CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (DIPPS) from *Archaeoglobus fulgidus*. *A. fulgidus* is a hyperthermophilic archaeon with an optimum growth temperature around 83°C. IPCT/DIPPS is involved in the synthesis of a compatible solute (di-*myo*-inositol-1,3´-phosphate), assumed to confer thermostability to cellular components and has only been found in (hyper)thermophilic organisms (Borges *et al*., 2010). It comprises two domains: a cytoplasmic N-terminal domain (IPCT) homologous to nucleotidyltransferases and a membrane C-terminal domain (DIPPS) that belongs to the CDP-alcohol phosphatidyltransferase family.

IPCT/DIPPS was heterologously overexpressed in *Escherichia coli*, solubilized with detergent, and successfully purified by immobilized

metal ion affinity chromatography. The 55 kDa protein was crystallized using either the vapor diffusion method or in lipidic mesophases. The mesophases approach yielded better quality crystals (best diffraction around 2.8 Å resolution at synchrotron sources) compared to the more traditional vapor diffusion method (diffraction *ca.* 4.5 Å). The IPCT/DIPPS structure was solved by molecular replacement using the X-ray structure of IPCT previously determined by Brito *et al.* (2012) and further completed by iterative cycles of model building and refinement. The structure reveals that the DIPPS domain consists mainly of 6 transmembrane helices and a helix connector to the IPCT domain, which binds weakly to the solvent exposed edge of the DIPPS domain. The crystal structure of IPCT/DIPPS is arranged as a dimer, with an interface formed by three transmembrane helices from each chain. A metal binding site is present in the DIPPS domain, where a putative magnesium ion is surrounded by 4 aspartate residues, 3 of which belong to the consensus sequence of CDP-alcohol phosphatidyltransferase family. Moreover, the substrate L-*myo*-inositol 1-phosphate was modeled into the catalytic site of the DIPPS domain. Mutagenesis studies allowed identification of several residues essential for DIPPS activity. Because CDP-*myo*-inositol, the product of IPCT activity, is a substrate of DIPPS, the arrangement of both domains in the X-ray structure is somewhat surprising, as their active sites are not facing each other. Interestingly, in most organisms IPCT and DIPPS are present as a single gene product, although separate genes are also found in

several other organisms (Gonçalves *et al*., 2012). It is yet unclear whether cooperation exists between the domains harboring these two consecutive activities. At present, it is also not known whether conformational changes occur upon ligand binding at the IPCT domain, which could eventually bring it into a more favorable orientation towards the active site of DIPPS.

DIPPS is a member of the CDP-alcohol phosphatidyltransferase class I family, for which no structure was yet available. This family comprises integral membrane proteins, such as choline/ethanolamine phosphotransferase and cardiolipin synthase. Members of this family are present in all domains of life and usually involved in phospholipid biosynthesis. The three-dimensional structure of IPCT/DIPPS determined at 2.8 Å resolution will have a very significant impact in the field of lipid biosynthesis and will certainly pave the way to other studies, e.g. mechanism of reactions, homology modeling followed by structure-based drug design.

The second enzyme described in this thesis is α-phosphoglucomutase from *Lactococcus lactis* (APGM). This enzyme plays an important role in carbohydrate metabolism by catalyzing the reversible conversion of α-glucose 1-phosphate to glucose 6-phosphate. Phosphoglucomutase is widespread in living organisms from bacteria to humans. It mediates utilization of phosphoglucose either for biosynthesis or, upon sugar source deprivation, for glycolysis. Thus, this enzyme is a crucial link between catabolic and anabolic processes (Neves *et al*., 2006).

The 30 kDa APGM was overexpressed in *Lactococcus lactis* and purified by immobilized metal ion affinity chromatography. It was crystallized by vapor diffusion method and X-ray diffraction data were collected at synchrotron source to 1.5 Å resolution. APGM structure was determined by single wavelength anomalous dispersion using a crystal soaked with a platinum compound. APGM structure shows that the most similar known structures are not the common bacterial α-phosphoglucomutases that belong to phosphohexomutase superfamily, but the eukaryotic phosphomannomutases (PMMs), though APGM is not able to utilize phosphomannose. APGM and PMMs belong to the haloacid dehalogenase superfamily (HADSF) and APGM is the first member of this superfamily with strict specificity for α-glucose 1-phosphate. APGM is present as a dimer in both the crystal structure and in solution, an arrangement that may be functionally relevant; however the dimer interface differs from that found in eukaryotic PMMs. Structural superposition of APGM and PMMs highlights their similarity and differences. Like other members of HADSF, APGM comprises two domains, the cap and core, which open to bind substrate and then close to provide a solvent-exclusive environment for catalysis (Silvaggi *et al*., 2006). Interestingly, the cap domain in APGM structure is found in a tightly closed conformation in comparison with other eukaryotic PMMs. Although we could not obtain the structure of APGM with ligands to assess the conformational changes induced upon their binding, in the APGM

crystal structure a sulfate and a glycerol were found at the enzyme active site. Both of these molecules partially mimic the substrate, which together with structural comparisons with other PMMs and docking studies provided insights into both the catalytic mechanism and the strict substrate selectivity of APGM.

RESUMO

As reacções químicas são essenciais para a sobrevivência dos organismos vivos e manutenção de inúmeras funções, desde a reprodução ao crescimento celular. A exposição destes organismos a mudanças contínuas no ambiente exercem uma pressão selectiva para uma rápida adaptação, que pode envolver a redução do crescimento face à privação da fonte de açúcar ou a minimização dos efeitos prejudiciais causados pelo stress térmico. Esta tese de doutoramento inclui a caracterização estrutural de duas enzimas envolvidas na resposta a alterações ambientais.

A primeira proteína de estudo no âmbito do trabalho de doutoramento é a enzima bifuncional citidililtransferase de L-*myo*-inositol 1-fosfato (IPCT) e fosfotransferase de CDP-L-*myo*-inositol:L-*myo*-inositol 1-fosfato (DIPPS) de *Archaeoglobus fulgidus*. *A. fulgidus* é um arqueão hipertermofílico com temperatura óptima de crescimento à volta de 83°C. A IPCT/DIPPS está envolvida na síntese de um soluto compatível (1,3´-di-myo-inositol-fosfato, DIP), que confere termoestabilidade aos componentes celulares dos organismos e só ainda foi possível encontrar esta enzima em (hiper)termófilos (Borges e al., 2010). A IPCT/DIPPS contém dois domínios: o N-terminal citoplasmático (IPCT) homólogo de nucleotidiltransferases e o C-terminal membranar (DIPPS) que pertence à família de fosfatidiltransferase de CDP-álcool.

A enzima IPCT/DIPPS foi sobre-expressa heterologamente em *Escherichia coli*, solubilizada com detergente e purificada com sucesso por cromatografia de afinidade do ião metálico imobilizado. A proteína de 55 kDa foi cristalizada usando o método da difusão do vapor ou em mesofases lipídicas. A técnica das mesofases provou ser mais adequada para o crescimento de cristais de melhor qualidade (a melhor diffracção obtida foi cerca de 2.8 Å em sinctrotrão) comparada com o método mais tradicional da difusão do vapor (diffracção aproximada de 4.5 Å). A estrutura de IPCT/DIPPS foi resolvida pelo método da substituição molecular, usando a estrutura de raios-X da IPCT determinada por Brito *e al.* (2012). De seguida, foram realizados ciclos iterativos de construção do modelo e refinamento cristalográfico. A estrutura revela que o domínio DIPPS é constituído essencialmente por seis hélices transmembranares e uma hélice conectora ao domínio IPCT, que liga fracamente à zona exposta ao solvente do domínio DIPPS. A estrutura cristalina de IPCT/DIPPS forma um dímero, sendo a sua interface composta por três hélices transmembranares de cada cadeia polipeptídica. No domínio DIPPS encontra-se um local de ligação a um metal, onde existe um magnésio putativo, rodeado por 4 aspartatos, 3 dos quais pertencem à sequência de consenso da família fosfatidiltransferase de CDP-álcool. O substrato L-*myo*-inositol 1-fosfato, foi modelado no centro activo do domínio DIPPS. Ensaios de mutagénese permitiram identificar diversos amino ácidos essenciais para a actividade catalítica da DIPPS. Sendo o produto da reacção da IPCT (CDP-L-*myo*-inositol)

substrato da DIPPS, o arranjo de ambos os domínios na estrutura cristalina não deixa de ser de algum modo surpreendente, dado que os seus centros activos não estão orientados um para o outro. É interessante notar que na maioria dos organismos, um único gene codifica para a proteína bifuncional IPCT/DIPPS, embora existam alguns organismos, onde os genes se encontram separados (Gonçalves *et al*., 2012). Não é contudo evidente se existe cooperação entre os domínios que abrigam estas duas actividades consecutivas. De momento, também não se sabe se ocorrem alterações conformacionais aquando da ligação do ligando ao domínio IPCT, que poderia eventualmente posicioná-lo numa orientação mais favorável relativamente ao centro catalítico da DIPPS.

A DIPPS é membro da família da classe I de fosfatidiltransferase de CDP-álcool, para a qual ainda não havia nenhuma estrutura caracterizada. Esta família inclui proteínas membranares integrais, tais como fosfotransferase de colina/etanolamina e sintase da cardiolipina. Os membros desta família estão presentes em todos os domínios da vida e encontram-se geralmente envolvidos na biosínese de fosfolípidos. A estrutura tri-dimensional da IPCT/DIPPS determinada a 2.8 Å de resolução terá um impacto muito significativo no campo da biosíntese de lípidos e irá certamente abrir portas a outros estudos, tais como, mecanismos das reacções, modelação molecular por homologia e desenvolvimento de fármacos baseado na estrutura (*rational drug design*).

A segunda enzima descrita nesta tese é α-fosfoglucomutase de *Lactis de Lactococcus* (APGM). Esta enzima desempenha um papel importante no metabolism dos hidrato de carbono catalisando a conversão reversível da α-glucose 1-fosfato a glucose 6-fosfato. A fosfoglucomutase está presente nos organismos vivos, das bactérias aos seres humanos. A APGM medeia a utilização da fosfoglucose para o biosíntese ou, em caso de privação da fonte do açúcar, para a glicólise. Assim, esta enzima faz a ligação entre os processos catabólicos e anabólicos (Neves *et al*., 2006).

A APGM (~30 kDa) foi sobre-expressa em *Lactococcus lactis* e purificada por cromatografia de afinidade do ião metálico imobilizado. Foi cristalizada pelo método da difusão do vapor e os dados de diffracção de raios-X foram recolhidos em instalações de sincrotrão a 1.5 Å de resolução. A estrutura da APGM foi determinada pelo método da dispersão anómala com um único comprimento de onda usando um cristal derivatizado com um composto de platina (Pt). A estrutura da APGM mostra que as estruturas que lhe são mais homólogas não são as α-fosfoglucomutases bacterianas que pertencem à superfamília das fosfohexomutases mas as fosfomanomutases eucarióticas (PMMs), embora a APGM não possa utilizar a fosfomanose como substrato. APGM e PMMs pertencem à superfamília *haloacid dehalogenase* (HADSF) e a APGM é o primeiro membro desta superfamília com estrita especificidade para a α-glucose 1-fosfato. A APGM encontra-se na forma dimérica na estrutura cristalina e em solução, arranjo este

que pode ser funcionalmente relevante. No entanto, a interface do dímero na APGM difere daquela encontrada nas PMMs eucarióticas. A sobreposição estrutural da APGM e das PMMs destaca as suas semelhanças e diferenças. Tal como os outros membros da HADSF, a APGM contém dois domínios, a tampa e o núcleo, que abrem para possibilitar a ligação do substrato e depois fecham, providenciando um ambiente favorável à catálise com exclusão de solvente (Silvaggi *et al*., 2006). O domínio da tampa da estrutura da APGM encontra-se numa conformação mais fechada em comparação com as restantes estruturas das PMMs. Embora não tenha sido possível obter a estrutura da APGM com ligandos para caracterizar as alterações conformacionais induzidas pelos mesmos, um sulfato e um glicerol estão presentes no centro activo da enzima. Ambas estas moléculas mimetizam parcialmente o substrato, e em conjunto com análise estrutural de outras PMMs e estudos de modelação fornecem dados importantes para uma melhor compreensão do mecanismo catalítico e da estrita selectividade da APGM.

ABBREVIATIONS

α-PGM	α-phosphoglucomutase
APGM	α-phosphoglucomutase from *Lactococcus lactis*
ATP	adenosine triphosphates
β-PGM	β–phosphoglucomutase
BLAST	basic local alignment search tool
CMP	cytidine monophosphate
CTP/CDP	cytidine triphosphate/diphosphate
CDP-OH_P_trans	CDP-alcohol phosphatidyltransferase family
CMC	critical micelle concentration
CSP	consensus sequence pattern of the CDP-OH_P_trans
CSS	complex formation significance score (in PISA server)
Cymal-5	5-cyclohexyl-1-pentyl-β-D-maltoside
DAG	diacylglycerol
DDM	*n*-dodecyl-β-D-maltoside
DIPPS	CDP-L-*myo*-inositol *myo*-inositolphosphotransferase
DIPP	di-*myo*-inositol-1,3´-phosphate-1´-phosphate
DIP	di-*myo*-inositol-1,3´-phosphate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
GDP	guanosine diphosphate
GPI	glycero-phospho-*myo*-inositol
HAD	haloacid dehalogenases
HADSF	haloacid dehalogenase superfamily
IPCT	CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani culture media
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance spectroscopy
OD	optical density
OG	n-octyl-β-D-glucoside
PCR	polymerase chain reaction
PDB	protein data bank
PGP	phosphatidylglycerophosphate
PMM	eukaryotic phosphomannomutases
PMSF	phenylmethanesulfonylfluoride

rmsd	root-mean-square deviation
rpm	revolutions per minute
SAD	single-wavelength anomalous diffraction
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
TB	terrific broth culture media
TCEP	tris(2-carboxyethyl)phosphine
TLS	translation libration screw-motion (method of structure refinement)
TM	transmembrane helix
Triton X-100	nonaethylene glycol octylphenol ether
UDP/UTP	uridine diphosphate/triphosphate
v/v	volume by volume
WT	wild-type enzyme
w/v	weight by volume

TABLE OF CONTENTS

Chapter 1 - INTRODUCTION

Chapter 2 - STRUCTURAL STUDY OF AN ATYPICAL α-PHOSPHOGLUCOMUTASE

Chapter 3 - STRUCTURAL STUDY OF MEMBRANE ENZYME IPCT/DIPPS

Chapter 1

INTRODUCTION

1.1 HALOACID DEHALOGENASE SUPERFAMILIY OF ENZYMES IN PHOSPHOTRANSFER REACTIONS

1.1.1 HAD SUPERFAMILIY

Haloacid dehalogenase (HAD) superfamily (CL0137), includes a diverse range of enzymes that use an aspartate carboxylate as nucleophile (Aravind et al., 1998; Koonin and Tatusov, 1994). There are ~79000 sequences classified into this superfamily in Pfam database and even more sequences, 148000, classified as HAD-like in the UniProt database (as of 19th February 2013). Members of this family are found in all domains of life and catalyze a variety of reactions with hydrolases being predominant in the superfamily (Figure 1.1), including the phosphoesterases, ATPases, phosphonatases, dehalogenases and sugar phosphomutases, which utilize a variety of different substrates (Burroughs et al., 2006). HAD superfamily is divided into 21 Pfam families, of which 13 are also represented in the InterPro database (Table 1.1).

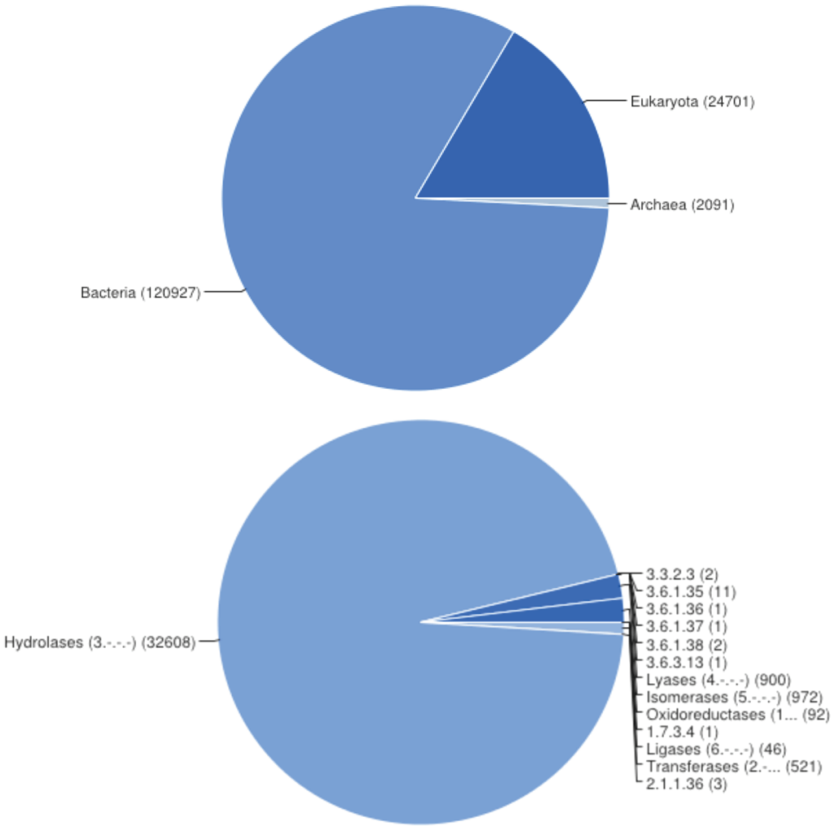


Figure 1.1 Taxonomy (top) and enzyme classes (bottom) of the haloacid dehalogenase superfamily sequences classified in UniProtKB. The enzyme classes: EC 3.3.2.3 - epoxide hydrolase, EC 3.6.1.35 - H^+-exporting ATPase, EC 3.6.1.36 - H^+/K^+-exchanging ATPase, EC 3.6.1.37 - Na^+/K^+-exchanging ATPase, EC 3.6.1.38 - Ca^{2+}-transporting ATPase, EC 3.6.3.13 - aminophospholipid-transporting ATPase, EC 1.7.3.4 - hydroxylamine oxidase, EC 2.1.1.36 - TRNA (adenine-N1-)-methyltransferase.

Table 1.1 Enzyme families classified into HAD in the Pfam and InterPro databases. The number of sequences refers to Pfam database on 13/03/2013.

Abbreviation	Description	Pfam	Sequences	InterPro
HAD_2	Haloacid dehalogenase-like hydrolase	PF13419	22576	
hydrolase	Haloacid dehalogenase-like hydrolase	PF00702	18348	IPR005834
hydrolase_3	haloacid dehalogenase-like hydrolase	PF08282	13347	IPR013200
HAD	haloacid dehalogenase-like hydrolase	PF12710	6179	
hydrolase_like	HAD-hydrolase-like	PF13242	5583	
hydrolase_6	Haloacid dehalogenase-like hydrolase	PF13344	3060	
NIF	NLI interacting factor-like phosphatase	PF03031	2044	
trehalose_PPase	Trehalose-phosphatase	PF02358	1750	IPR003337
hydrolase_like2	Putative hydrolase of sodium-potassium ATPase alpha subunit	PF13246	1296	
acid_phosphat_B	HAD superfamily, subfamily IIIB, acid phosphatase	PF03767	1117	IPR005519
PNK3P	Polynucleotide kinase 3 phosphatase	PF08645	685	IPR013954
S6PP	Sucrose-6F-phosphate phosphohydrolase	PF05116	588	IPR006380
PMM	Eukaryotic phosphomannomutase	PF03332	465	IPR005002
NT5C	5' nucleotidase, deoxy(Pyrimidine), cytosolic type C protein	PF06941	458	IPR010708
LNS2	Lipin/Ned1/Smp2	PF08235	430	IPR013209
5_nucleotid	5' nucleotidase family	PF05761	371	IPR008380
PGP_phosphatase	Mitochondrial PGP phosphatase	PF09419	250	IPR010021
Acid_PPase	Acid phosphatase	PF12689	225	
Put_phosphatase	Putative phosphatase	PF06888	213	IPR006384
UMPH-1	Pyrimidine 5'-nucleotidase	PF05822	159	
DUF705	Protein of unknown function	PF05152	68	IPR007827

RANGE OF ARCHITECTURES

Structural characterization of proteins from the HAD superfamily resulted in over 400 entries in the Protein Data Bank (PDB) as indicated in the Pfam database. Members of this superfamily comprise a characteristic HAD domain with a Rossmannoid fold and with a catalytic aspartate located at the end of the first β-strand of the central core β-sheet (Burroughs et al., 2006). The structural element characteristic of this domain includes an almost completed one helical turn, just after the first β-strand and followed by a β-hairpin turn (Figure 1.2).

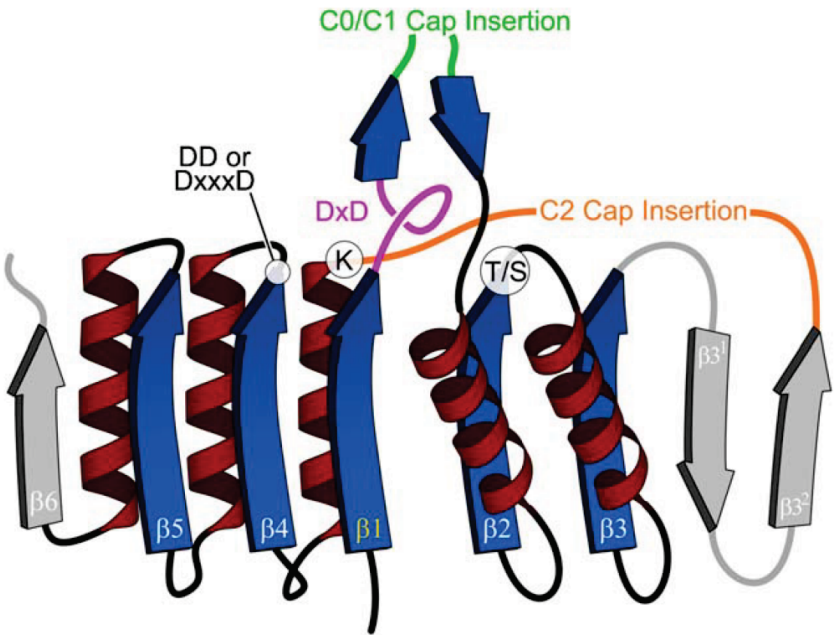


Figure 1.2 Topology diagram of the HAD domain. Indicated are insertions of the C0/C1 and C2 structural classes. In pink is shown the characteristic nearly completed helical turn and in green the β-hairpin. This figure is taken from (Allen and Dunaway-Mariano, 2009).

HAD domain (core domain) provides a catalytic scaffold for reactions with conserved amino acid residues, including the catalytic aspartate that is transiently phosphorylated in the phosphotransfer enzymes, and the residues binding magnesium ion (Allen and Dunaway-Mariano, 2009).

The adaptation to a wide range of functions and substrates in the HAD superfamily enzymes is enabled by additional insertions into the conserved HAD domain (Figure 1.2). Based on the topology of insertion and its position, the superfamily can be divided into structural classes C0, C1, C2 (Allen and Dunaway-Mariano, 2009; Burroughs et al., 2006).

The C0 class contains the shortest insert and thus can be considered as the most basic member of HAD superfamily. This class contains mainly phosphatases showing different degrees of elaboration concerning the insert in the β-hairpin turn; none in deoxy-D-mannose-octulosonate 8-phosphate phosphatase (e.g. PDB code: 1j8d) to one additional β-strand in CTD phosphatases (e.g. PDB code: 3pgl).

The C1 class contains a larger insert in the characteristic β-hairpin (Figure 1.2), which folds into a cap domain separated from the HAD domain (core domain). The cap domains in this class may be only α-helical as in β-phosphoglucomutases (e.g. PDB code: 1lvh) or show α+β fold, as P-type ATPases (e.g. PDB code: 1t5t).

The C2 class possesses a cap domain inserted into the loop after β3-strand. Caps of this class are α+β and two major types can be

distinguished. One resembles the cap of Cof-type phosphatases and the second resembles the cap of NagD-like phosphatases. The Cof-type phosphatases contain a central three stranded antiparallel β-sheet (PDB code: 1l6r), which can be extended by a number of additional secondary structure elements (PDB code: 1xvi). Moreover, in the NagD-like phosphatases, the basic structural element of the cap domain consists of four-stranded parallel β-sheet, along with one N-terminal antiparallel strand; the β-hairpin is positioned after the third strand and this fold can be observed in 4-nitrophenylphosphatase (PDB code: 1VJR).

CATALYTIC MECHANISM OF PHOSPHOTRANSFER

The common core domain of the HAD superfamily provides the conserved catalytic residues and shows similar reaction mechanisms (Figure 1.3). This is particularly evident in the case of phosphotransferases, as they all utilize a strictly conserved aspartate as a nucleophile forming an intermediate aspartylphosphate. In the subsequent stage, depending on the type of reaction, the intermediate is either hydrolyzed as in phosphatases or phosphoryl is transferred to a hydroxyl group, as in phosphohexomutases (Figure 1.3).

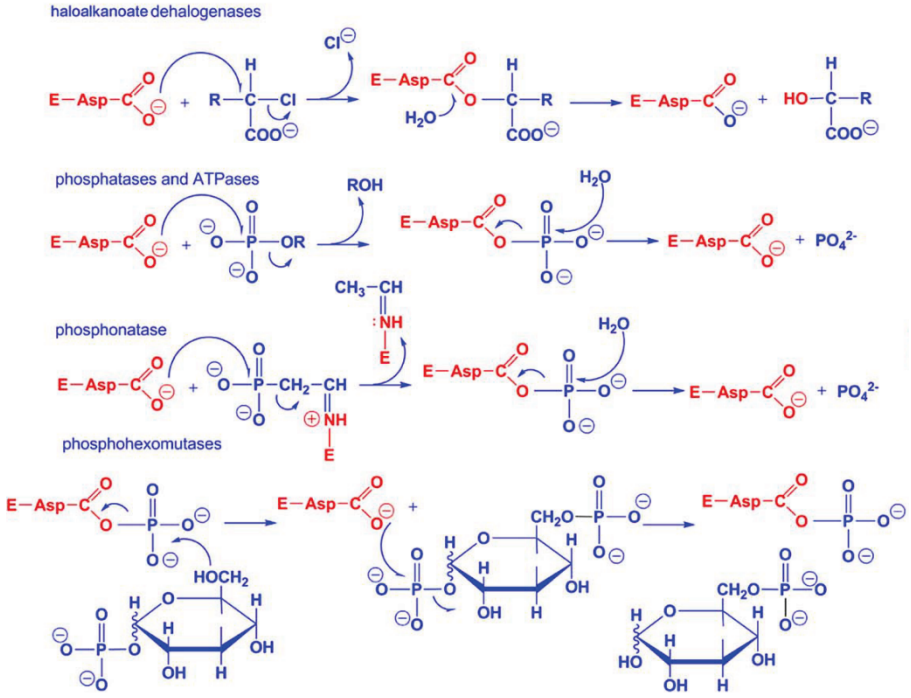


Figure 1.3 Reaction mechanisms utilized within the HAD superfamily. This figure is taken from Burroughs et al., 2006.

Besides the nucleophilic aspartate sited in loop 1 and coordinated to Mg^{2+}, there is a number of other residues located in the catalytic site of the core domain, which are conserved (Figure 1.4). In loop 1 following the first β-strand and positioned two residues after the nucleophilic Asp, is another aspartate with acid/base function. It protonates the leaving group in the first stage of the reaction and acts as proton acceptor in the final stage of phosphoryl transfer from aspartylphosphate.

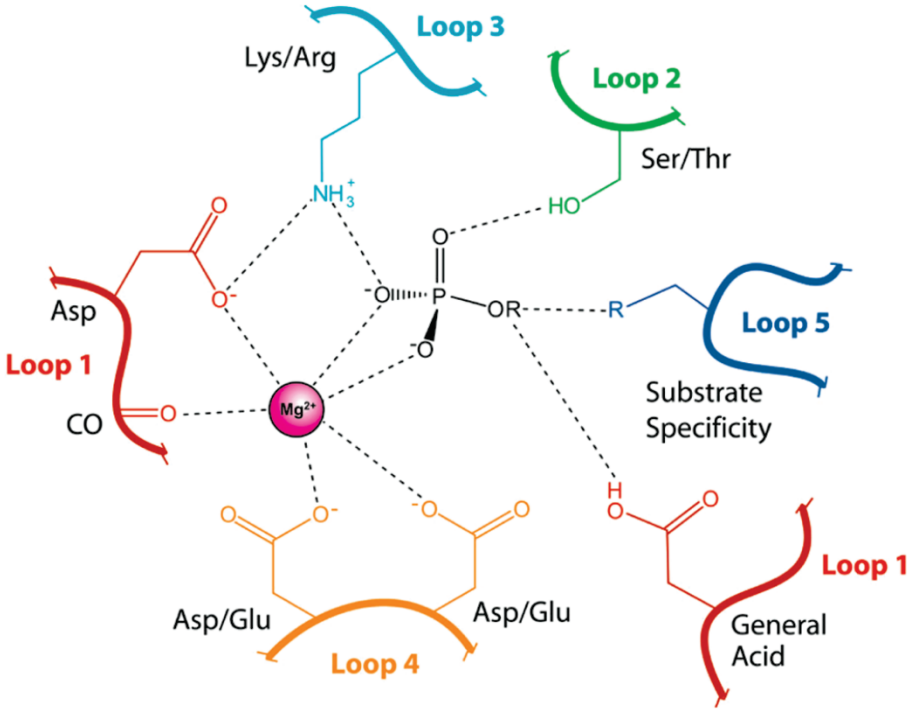


Figure 1.4 Schematic representation of the HAD catalytic scaffold with indicated interactions between phosphoester substrate and conserved residues. This figure is taken from Dai et al., 2009.

In phosphotransferases, the metal ion is coordinated to three negatively charged Asp/Glu residues and the phosphoryl group of the substrate, thus playing a role in their electrostatic stabilization and probably also stabilizing the transition state by neutralizing its charge (Allen and Dunaway-Mariano, 2009).

1.1.2 EUKARYOTIC PHOSPHOMANNOMUTASES AND β-PHOPHOGLUCOMUTASES

Eukaryotic phosphomannomutases (PMM, IPR005002) resemble the architecture of Cof-type phosphatase of the C2 class of HAD superfamily (Figure 1.5a). These enzymes are involved in the reversible interconversion of mannose 6-phosphate and mannose 1-phosphate, which is required in the synthesis of GDP-mannose needed in the glycosylation of human proteins (Schollen et al., 1998; Tomavo et al., 1992). Moreover, PMMs also catalyze the interconversion of glucose 1-phosphate to glucose 6-phosphate (Silvaggi et al., 2006).

The two interconversion reactions of PMMs can also be catalyzed by unrelated enzymes, members of different α-D-phosphohexomutase superfamily widely found in bacteria and in other domains of life. However, as shown in the Figure 1.5c, the protein topology of α-D-phosphohexomutases is quite different from the HAD superfamily (Figure 1.5a-b). In addition, the α-D-phosphohexomutase superfamily utilizes serine as active site residue involved in phosphorylation, thus the reaction mechanism is also distinct from the eukaryotic phosphomannomutases (Shackelford et al., 2004).

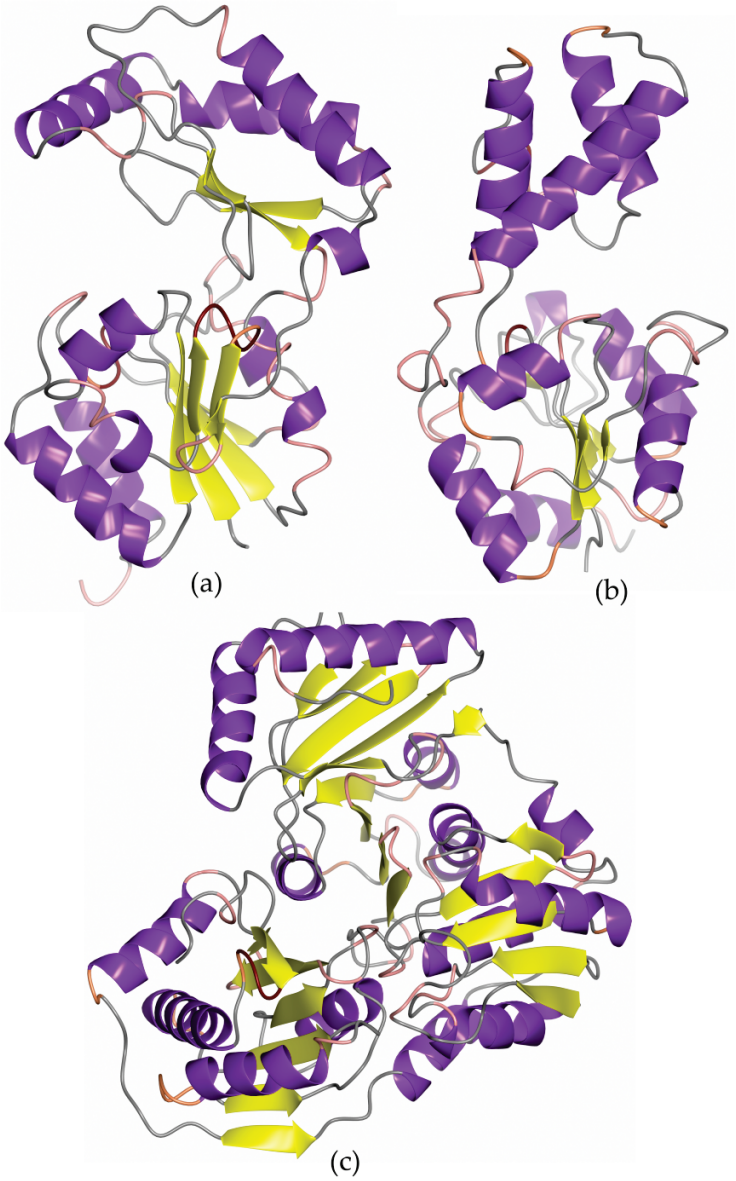


Figure 1.5 Representation of different folds of HAD superfamily members: (a) the human phosphomannomutase 1 (PDB code: 2fue), (b) β-phosphoglucomutase from *Lactococcus lactis* (PDB code: 1lvh); and α-D-phosphohexomutase superfamily member: (c) phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa* (PDB code: 2h4l). Figure drawn with CCP4MG (McNicholas et al., 2011).

HAD superfamily also includes another enzyme, β-phosphoglucomutase, which catalyzes the reversible conversion of a different phosphoglucose anomer, β-glucose 1-phosphate to β-glucose 6-phosphate. Interestingly, the recognition of this anomeric substrate was achieved with a very different topology of the cap domain from the one of eukaryotic phosphomannomutases (Figure 1.5b). The cap insertion into the β-hairpin following the first β-strand categorizes the β-phosphoglucomutase into the class C1 of the HAD superfamily.

Eukaryotic PMMs share with β-phosphoglucomutases the requirement for an activator of the reaction, respectively α or β anomer of glucose-1,6-bisphosphate (also α-mannose-1,6-bisphosphate in PMMs) (Silvaggi et al., 2006; Zhang et al., 2005). These activators are utilized to phosphorylate the nucleophile Asp yielding an active enzyme, which in further cycles utilize monophosphorylated substrates (Figure 1.6). For the well studied β-phosphoglucomutase from *Lactococcus lactis* (Griffin et al., 2012; Lahiri et al., 2002; Zhang et al., 2005) phosphoryl transfer was proposed from aspartyl of the activated enzyme to the substrate forming an intermediate β-glucose-1,6-bisphosphate intermediate. This intermediate needs to be rearranged, likely by its release from the active site followed by recapturing in a different position with the phosphoryl of the substrate to be removed located next to the nucleophile Asp. In the final stage of the reaction the active form of the enzyme with aspartylphosphate is regenerated and the product

released. It is also suggested that when the reaction takes place the cap domain closes tightly over the core domain to exclude the solvent from the active site and prevent transfer of the phosphoryl to water molecule (Dai et al., 2009). However, for the release and rebinding of the intermediate in the rearranged position, the cap needs to open, since the space available in the active site is not sufficient for the flip to take place. Similar mechanism was also suggested for human PMM 1 (Silvaggi et al., 2006).

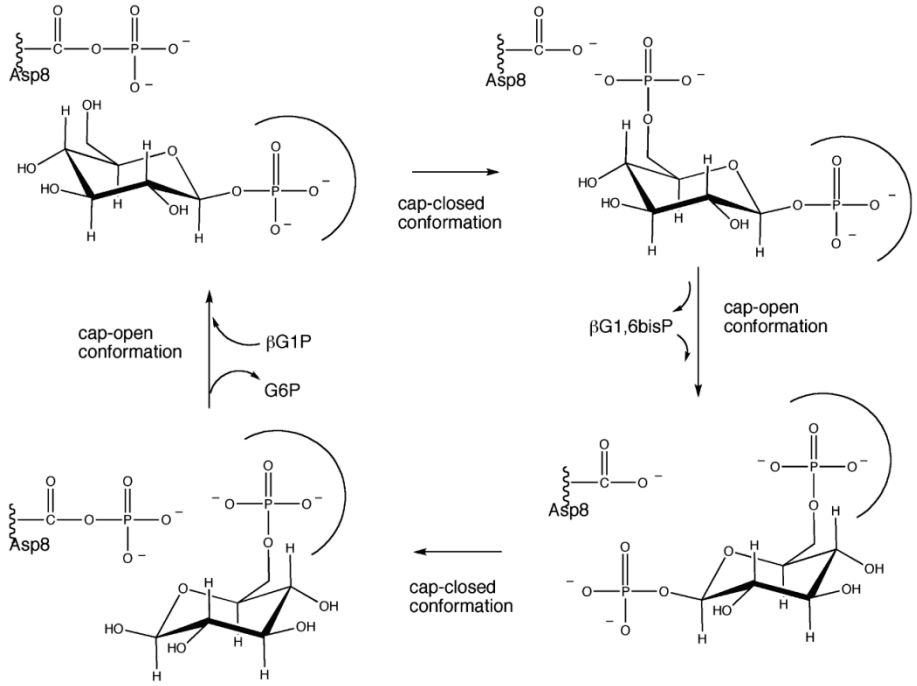


Figure 1.6 Catalytic cycling in β-phosphoglucomutase from *Lactococcus lactis*. This figure is taken from Dai et al., 2009.

1.1.3 α-PHOSPHOGLUCOMUTASE FROM *LACTOCOCCUS LACTIS*

Bacterial α-phosphoglucomutases studied so far belong to the α-D-phosphohexomutase superfamily. In 2006, Neves et al. identified and characterized a gene of a bacterial α-phosphoglucomutase, which was unrelated to α-D-phosphohexomutase superfamily (Neves et al., 2006). This gene of *Lactococcus lactis* with UniProt accession number Q00G41 encodes for a protein with 252 amino acids. In fact *Lactococcus lactis* does not possess a gene commonly found in bacteria encoding for a protein of about 400 amino acids with α-phosphoglucomutase activity of the α-D-phosphohexomutase superfamily. Instead, the newly identified gene provides a replacement for this.

Sequence analysis revealed the highest sequence identity to characterized eukaryotic phosphomannomutases of about 24% and only 10% to β-phosphoglucomutase. Moreover, the conserved motifs of the HAD superfamily were also found in the protein sequence, in particular the catalytic nucleophile Asp8 and the acid/base Asp10.

Functional characterization of this enzyme revealed that it does not utilize β-glucose 1-phosphate or α-mannose 1-phosphate (Neves et al., 2006). However, this lactococcal α-phosphoglucomutase catalyzes reversible conversion of α-glucose 1-phosphate to glucose 6-phosphate. From now on this α-phosphoglucomutase from *Lactococcus lactis* will be abbreviated as APGM.

Impact of the APGM gene expression was studied *in vivo* in *Lactococcus lactis*. It was shown that the expression of this gene was needed for viability of *Lactococcus lactis*. Lowered levels of APGM activity were sufficient for cell growth but this was accompanied with changes in the cell morphology and cell division (Neves et al., 2006). It was therefore concluded that the synthesis of the UDP-glucose, precursor for cell wall biosynthesis, relies on APGM activity. Moreover, overexpression of APGM in *L. lactis* grown on galactose indicated that this enzyme activity is the bottleneck in galactose metabolism. On the other hand, in *L. lactis* grown on glucose the APGM mainly provides substrates for biosynthetic pathways (Neves et al., 2006).

It was shown that similar to other members of the HAD superfamily APGM requires presence of Mg^{2+} for activity. Also α-glucose-1,6-bisphosphate activator is needed to initiate the catalysis. The reaction can proceed into both directions, though the V_{max} in the conversion of glucose 6-phosphate to α-glucose 1-phosphate is four times lower than in the other direction. Also the K_m was determined for glucose 6-phosphate in the millimolar range, while for α-glucose 1-phosphate it was 71 μM (Neves et al., 2006).

The similarities in sequence and activity between lactococcal APGM and eukaryotic PMMs raise an obvious question about the structural relation between these enzymes. Moreover, the strict specificity of APGM towards α-glucose 1-phosphate and not its stereoisomer α-mannose 1-phosphate like in the case of PMMs points to an

interesting subject of substrate specificity, which may rely on some subtle structural differences of the enzyme. In order to provide an explanation for the observed substrate differentiation we set off to determine an atomic model of APGM in this work. The crystallization and structure of APGM are described in Chapter 2.

1.2 CDP-OH PHOSPHATIDYLTRANSFERASE FAMILY & COMPATIBLE SOLUTES

1.2.1 CDP-OH PHOSPHATIDYLINOSITOL TRANSFERASE

CDP-alcohol phosphatidyltransferase family (IPR000462, PF01066) comprises enzymes that catalyze the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond. The majority of the enzymes characterized so far from this family are involved in phospholipids biosynthesis and share a conserved sequence pattern $DG(x)_2AR(x)_8G(x)_3D(x)_3D$ (PS00379). There are 8035 sequences classified into the family in the Pfam database and 12756 sequences in the UniProtKB (as of 3rd February 2013). Members of this family are found in all domains of life (Figure 1.7) and are typically 150 – 400 amino acid residues long.

Proteins of the CDP-alcohol phosphatidyltransferase (CDP-OH_P_trans) family are predicted to be located in the membrane. Transmembrane helices (TM) prediction performed with TMHMM Server v. 2.0 for 106 sequences of the family (the reviewed sequences of UniProtKB classified into the family; as on 3rd February 2013) indicates two to ten putative transmembrane helices (data not shown).

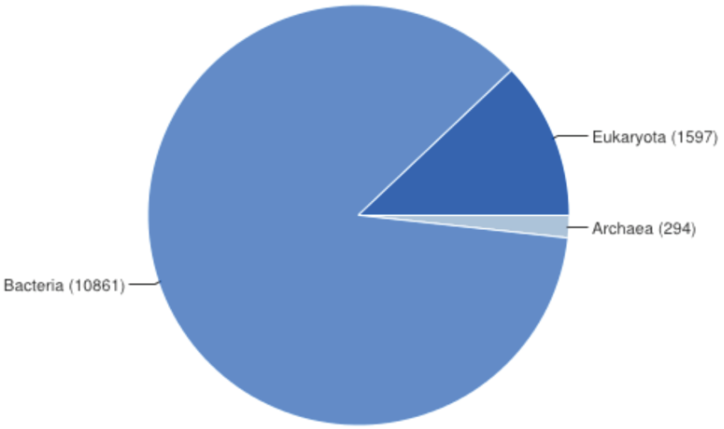


Figure 1.7 Taxonomy of the CDP-alcohol phosphatidyltransferase family sequences (number of sequences in brackets) classified in UniProtKB.

RANGE OF ACTIVITIES

The enzymes of CDP-alcohol phosphatidyltransferase family utilize a variety of different substrates and the common type of reaction involves the general reaction scheme:

$$\text{CDP–alcohol}_{(1)} + \text{alcohol}_{(2)} = \text{alcohol}_{(1)} – \text{P–alcohol}_{(2)} + \text{CMP} + \text{H}^+$$

However, not all enzymes involved in this type of reaction belong to CDP-alcohol phosphatidyltransferase family. A distinct family of peripheral membrane proteins, such as phospholipase D, phosphatidylserine synthase type (IPR016270, PF00614 or CDP-alcohol phosphatidyltransferase class-II family in the UniProt classification) can replace CDP-alcohol phosphatidyltransferase family enzymes in various organisms.

Depending on the substrate utilized by the enzyme, five subfamilies of CDP-alcohol phosphatidyltransferase family are already differentiated in the InterPro database (Hunter et al., 2012):

- CDP-diacylglycerol **glycerol**-3-phosphate 3-phosphatidyltransferase (IPR004570)
- CDP-diacylglycerol **serine** O-phosphatidyltransferase (IPR004533)
- CDP-diacylglycerol **inositol** 3-phosphatidyltransferase (IPR014387)
- **Choline**/**ethanolamine** phosphotransferase (IPR014472)
- Phosphatidyl**choline** synthase (IPR026027)

Another group of enzymes described in this thesis should also be included into the family:

- CDP-L-***myo*-inositol** ***myo*-inositol**phosphotransferase classified with enzyme entry EC 2.7.8.34

As it is apparent from the names of the subfamilies, the majority of the identified CDP-alcohol phosphatidyltransferase family enzymes are involved in phospholipid biosynthesis. Their presence in the eukaryotic lipid biosynthetic pathways is indicated in the Figure 1.8.
There are two routes providing substrates for the lipid biosynthesis by CDP-alcohol phosphatidyltransferases and in both of them one of the substrates needs to be activated with CDP for reaction (Lykidis, 2007). In the one initiated by CDP-DAG synthase, the CDP activated compound is diacylglycerol and the specific hydrophilic head group of the phospholipid is introduced via direct reaction with glycerol-3-phosphate (phosphatidylglycerol phosphate synthase), inositol

(phosphatidylinositol synthase) , serine (phosphatidylserine synthase) or phosphatidylglycerol (cardiolipin synthase). The second route utilizes diacylglycerol and the hydrophilic head group choline (CDP-choline:diacylglycerol phosphotransferase) or ethanolamine (CDP-ethanolamine:diacylglycerol phosphotransferase) is activated with CDP for reaction.

In bacteria, a different route of synthesis of phosphatidylcholine can also be found, which involves CDP-diacylglycerol and choline (phosphatidylcholine synthase) (Sohlenkamp et al., 2000). In archaea a specific type of phosphatidylserine synthase, archaetidylserine synthase was identified (Morii and Koga, 2003). This enzyme is involved in synthesis of ether lipids, very characteristic for archaea. Substrates for the reaction besides serine can be CDP-unsaturated archaeol with ether-linked geranylgeranyl chains and CDP-saturated archaeol with ether-linked phytanyl chains.

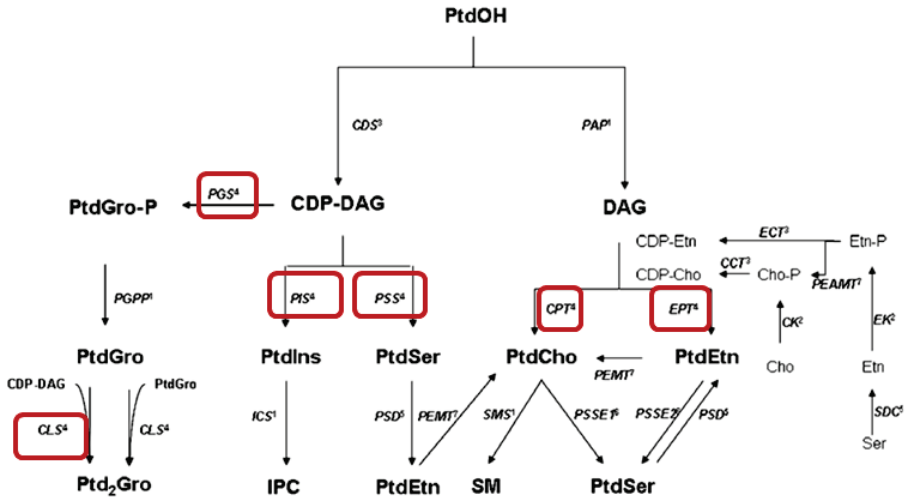


Figure 1.8 Lipid biosynthetic pathways in eukaryota. Enzymes belonging to CDP-alcohol phosphatidyltransferase family are indicated with red box. Substrates/products abbreviations: DAG – diacylglycerol, CDP-DAG - CDP-diacylglycerol, PtdOH - phosphatidic acid, PtdCho - phosphatidylcholine, PtdEtn - phosphatidylethanolamine, PtdIns - phosphatidylinositol, PtdSer - phosphatidylserine, PtdGro - phosphatidylglycerol, Ptd2Gro - cardiolipin, SM - sphingomyelin, IPC – inositolphosphorylceramide. Enyzme abbreviations: CCT - phosphocholine cytidylyltransferase, CDS - CDP-DAG synthase, CK - choline kinase, CLS - cardiolipin synthase, CPT - CDP-choline:diacylglycerol phosphotransferase, ECT - phosphoethanolamine cytidylyltransferase, EK - ethanolamine kinase, EPT - CDP-ethanolamine:diacylglycerol phosphotransferase, ICS - inositol phosphorylceramide synthase, PAP - phosphatidate phosphatase, PEMT - phosphatidylethanolamine methyltransferase, PEAMT - phosphoethanolamine methyltransferase, PIS - phosphatidylinositol synthase, PGPP - phosphatidylglycerol phosphate phosphatase, PGS - phosphatidylglycerol phosphate synthase, PSS - phosphatidylserine synthase, PSD - phosphatidylserine decarboxylase, PSSE - phosphatidylserine synthase via base-exchange, SMS - sphingomyelin synthase, SDC - serine decarboxylase. This figure is adapted from Lykidis, 2007.

However, not all of the CDP-alcohol phosphatidyltransferase enzymes are involved in synthesis of lipids. A compatible solute, di-*myo*-inositol-1,1′-phosphate (DIP) is found exclusively in (hyper)thermophilic archaea and bacteria (Gonçalves et al., 2012). It is assumed to confer thermostability to cellular components (Borges et al., 2010). DIP synthesis involves CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (DIPPS), a member of the CDP-alcohol phosphatidyltransferase family. It utilizes as substrates L-*myo*-inositol 1-phosphate and CDP activated L-*myo*-inositol 1-phosphate. Interestingly, it is not the only example of use of L-*myo*-inositol 1-phosphate by CDP-alcohol phosphatidyltransferases. It was reported that phosphatidylinositol in *Mycobacterium tuberculosis* was synthesized via reaction of CDP-diacylglycerol with L-*myo*-inositol 1-phosphate catalyzed by phosphatidylinositol phosphate synthase (Morii et al., 2010).

BIOCHEMICAL STUDIES

The strong hydrophobic character of CDP-alcohol phosphatidyltransferases, which are integral membrane proteins, makes difficult much of the biochemical studies. However, high interest in the biosynthetic pathways of phospholipids, one of the major components of the cell, drives steady progress in the field.

Breakthrough in the studies on CDP-alcohol phosphatidyltransferase enzymes was made in 1956 by the identification of the requirement

for the CDP activation of choline and ethanolamine in lipid synthesis (Kennedy and Weiss, 1956).

The first reports on purification of enzymes from this family from natural sources comes from 1976 when the phosphatidylglycerophosphate synthase from *Escherichia coli* and from *Bacillus licheniformis* were purified by substrate affinity chromatography (Dowhan, 2013; Hirabayashi et al., 1976; Larson et al., 1976). The proteins were solubilized with Triton X-100 (in the presence of glycerol) and CDP-diglyceride linked to sepharose was used in the crucial step of purification. A relatively high thermal stability for the *E. coli* protein was reported with the enzyme retaining its full activity after 5 min at 55 °C. Later, the pgsA gene of *E. coli* was cloned (Ohta et al., 1981), sequenced and confirmed to encode for phosphatidylglycerophosphate synthase *in vitro* (Gopalakrishnan et al., 1986). Similar purification procedures as for phosphatidylglycerophosphate were successfully used for phosphatidylserine synthase (Bae-Lee and Carman, 1984) and phosphatidylinositol synthase from *Saccharomyces cerevisiae* (Fischl and Carman, 1983). The latter purified enzyme was reconstituted into phospholipid vesicles and characterized in the absence of detergent (Fischl et al., 1986). Phosphatidylserine synthase from *Saccharomyces cerevisiae* was used to investigate the stereochemistry of the reaction using substrates stereospecifically labeled with O^{17} and O^{18} (Raetz et al., 1987). The reaction proceeds with inversion of configuration on the phosphorus, which suggests a single-displacement reaction

mechanism without formation of enzyme-bound intermediate (Figure 1.9).

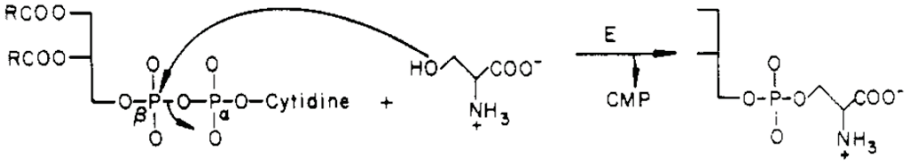


Figure 1.9 Proposed reaction mechanism of phosphatidylserine synthase. This figure is taken from Raetz et al., 1987.

The genes coding for yeast cholinephosphotransferase and ethanolaminephosphotransferase were sequenced in 1990 and 1991 (Hjelmstad and Bell, 1991, 1990). The catalytic activity of these enzymes was shown to depend on the presence of divalent ions with Mg^{2+} being the most activating ion followed by Mn^{2+} and Co^{2+} (McMaster et al., 1996).

In 1998, mutagenesis of amino acid residues within the family consensus sequence pattern, $DG(x)_2AR(x)_8G(x)_3D(x)_3D$ in cholinephosphotransferase from *Saccharomyces cerevisiae* was performed. Mutations to alanine or glycine inactivated the enzyme in four point mutations of the 2nd, 5th, 6th and 7th residues of the sequence pattern (Williams and McMaster, 1998).

A year later, the human choline/ethanolaminephosphotransferase was sequenced, cloned and expressed in yeast (Henneberry and McMaster, 1999). However, till today the choline- and

ethanolaminephosphotransferase have not been purified to homogeneity despite the variety of studies reported *in vivo* and partially purified cell extracts (Fagone and Jackowski, 2013).

The above description underlines the major steps in the progress of biochemical studies on enzymes of the CDP-alcohol phosphatidyltransferase family. Further studies on the family enzymes from other organisms followed. Two of the major obstacles mentioned in the literature include purification to homogeneity and the need for detergents to extract and solubilize integral membrane proteins. These obstacles contributed to the lack of detailed information on the structure of any member of the CDP-alcohol phosphatidyltransferase family. The increasing data on the impact of point mutations on the enzyme´s activity (Cutler et al., 2003; Solís-Oviedo et al., 2012) urge for atomic structure determination that would enable interpretation of these experiments and lead to a better understanding of the reaction mechanism.

1.2.2 COMPATIBLE SOLUTES AND IPCT/DIPPS

The enzymes of CDP-alcohol phosphatidyltransferase family have the requirement for a CDP activated substrate. Cytidylyltransferases of the nucleotidyltransferase family (PF00483) catalyze this activation for a variety of substrates. The CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (DIPPS; alternatively named di-*myo*-

inositol phosphate phosphate synthase) from *Archaeoglobus fulgidus* and a number of other (hyper)thermophiles (Gonçalves et al., 2012) can be found fused with CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT), the enzyme that activates its substrate. This bifunctional enzyme is involved in the biosynthesis of a compatible solute di-*myo*-inositol-1,3′-phosphate (DIP).

COMPATIBLE SOLUTES

Compatible solutes are defined as small organic molecules that do not interfere with the cell function (Brown, 1976). Accumulation of compatible solutes is linked with adaptation of the organisms to extreme saline environment and high temperatures. Organisms with optimal growth temperature between 60°C and 80°C (thermophiles) and above 80°C (hyperthermophiles) utilize some specific compatible solutes (Empadinhas and Da Costa, 2006). Those often found in (hyper)thermophiles include glutamate, aspartate, trehalose, mannosylglycerate, glucosylglycerate, di-*myo*-inositol-phosphate, diglycerol phosphate, di-mannosyl-di-*myo*-inositol-phosphate, di-*myo*-inositol-1,3′-phosphate (DIP). The last one, DIP, was found to accumulate rapidly in the cell upon supraoptimal temperatures (Gonçalves et al., 2003; Martins and Santos, 1995). The presence of DIP restricted to (hyper)thermopiles suggests a role in thermoadaptation (Gonçalves et al., 2003), further corroborated by protein stabilization studies *in vitro* (Santos et al., 2007).

IPCT/DIPPS

In 2007, the biosynthetic pathway of DIP was described and the genes of the enzymes involved were identified (Rodionov et al., 2007; Rodrigues et al., 2007). The first two steps of the reaction are catalyzed by the bifunctional enzyme IPCT/DIPPS (Figure 1.10). This enzyme consists of a cytoplasmic domain, IPCT, which activates L-*myo*-inositol-1-phosphate with CTP to CDP-L-*myo*-inositol and releases pyrophosphate. The consecutive reaction catalyzed by the DIPPS domain couples CDP-L-*myo*-inositol with another molecule of L-*myo*-inositol-1-phosphate to form di-*myo*-inositol-1,3′-phosphate-1′-phosphate. DIPPS, like other members of the CDP-alcohol phosphatidyltransferase family contains several transmembrane helices. The last step of the synthesis involves phosphate cleavage by a phosphatase and provides the compatible solute DIP.

In the same enzyme, DIPPS was also shown to utilize CDP-glycerol as a substrate, in addition to L-*myo*-inositol 1-phosphate, leading to synthesis of a different compatible solute, glycero-phospho-*myo*-inositol (GPI) (Rodrigues et al., 2007). GPI appears to be produced in response to combined osmotic and thermal stresses (Borges et al., 2006).

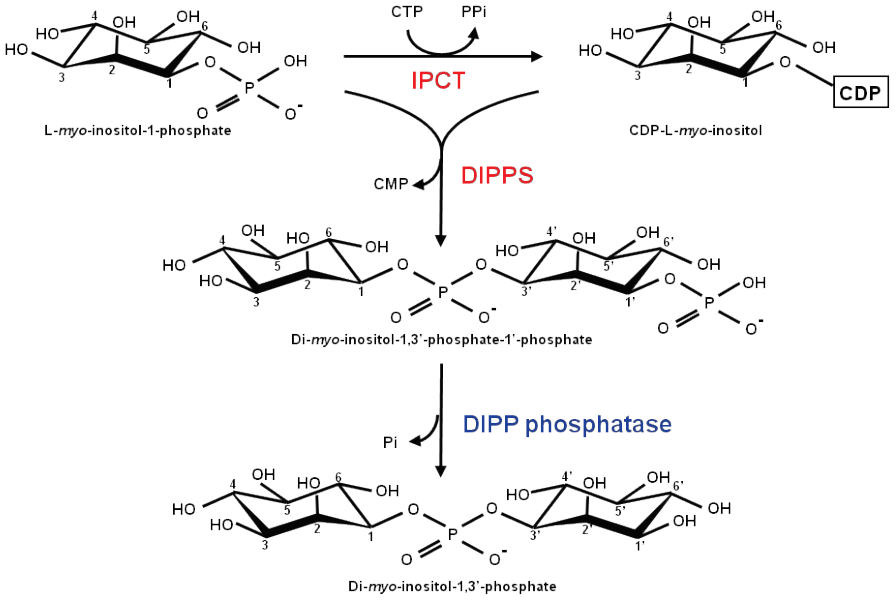


Figure 1.10 Biosynthetic pathway of di-*myo*-inositol-1,3′-phosphate. Abbreviations: IPCT - CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase, DIPPS - CDP-L-*myo*-inositol *myo*-inositolphosphotransferase, DIPP phosphatase - di-*myo*-inositol-1,3′-phosphate-1′-phosphate phosphatase. This figure is adapted from Rodrigues et al., 2007.

The genes encoding for IPCT and DIPPS were found in 33 (hyper)thermophiles within Archaea and Bacteria (Gonçalves et al., 2012). Several of them encode the two enzymes in a separate gene products, however the majority of the identified genes encode for a bifunctional enzymes IPCT/DIPPS, where the genes were fused. It was also proposed that DIPPS evolved from phosphatidylinositol phosphate synthase of the phospholipid biosynthetic pathway (Gonçalves et al., 2012).

X-RAY STRUCTURE OF IPCT FROM *ARCHAEOGLOBUS FULGIDUS*

The purification of the full-length bifunctional enzyme IPCT/DIPPS from *Archaeoglobus fulgidus* was unsuccessful (Rodrigues et al., 2007). Thus, the cytoplasmic IPCT domain alone has been expressed, purified, characterized and its crystallographic structure was solved (Brito et al., 2011). The IPCT activity was shown to be dependent on the presence of Mg^{2+}. Interestingly, the products of IPCT were undetectable below 60°C and the maximum activity was observed at 90 – 95°C. This optimal temperature for the enzyme activity is 10°C above the optimal growth temperature of *A. fulgidus*. In the biosynthetic pathway of phosphatidylcholine, the CDP-choline synthesis by a cytidylyltransferase was shown to be the bottleneck in the phospholipid synthesis (Vance, 2002). It is not known if the same applies to the synthesis of DIP, however, the observed optimum temperature for IPCT activity is consistent with the increase of the cellular DIP concentration at supraoptimal temperatures for *A. fulgidus* (Gonçalves et al., 2003).

The structure of IPCT from *A. fulgidus* was solved at 1.9 Å resolution (Figure 1.11; Brito et al., 2011). Two subdomains have been defined. The core subdomain has a seven-stranded β-sheet in the central part surrounded by 6 helices, a similar fold to the dinucleotide-binding Rossmann fold. The sugar binding subdomain consists of three-stranded short β-sheet. A pocket between the two subdomains forms the active site, which includes the signature sequence of

nucleotidyltransferases, G-X-T-R-$(X)_4$-P-K, and binds a citrate molecule in the crystal structure (Figure 1.11). Based on the comparison with homologous structures obtained with substrates, the IPCT substrates were fitted into the active site. Analysis of the CTP-fitted model provides clues regarding the strict substrate specificity for the CTP over other nucleosides. It was proposed that a tight loop (His89 - Gly95, numbering following Brito et al., 2011) excludes binding of purines and an additional H-bond formed between NH_2 amino group of cytidine and the main chain oxygen of Pro92 favors CTP over UTP binding.

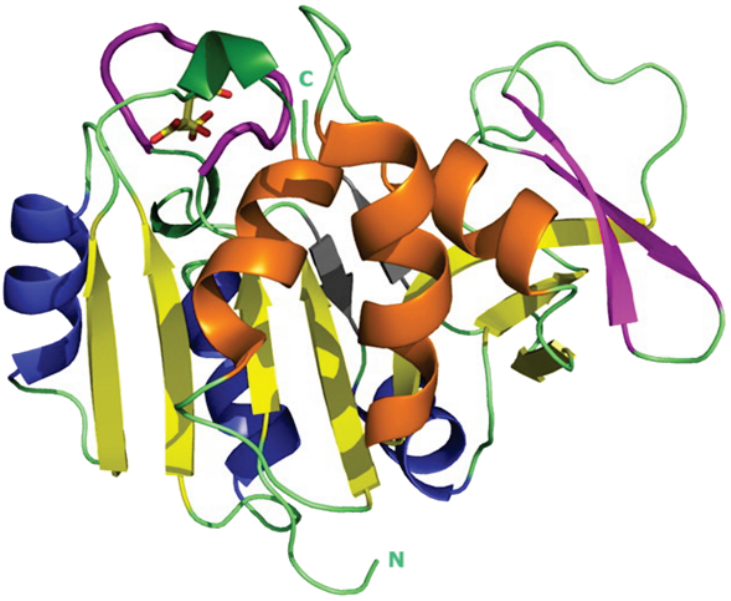


Figure 1.11 IPCT structure with citrate bound in the active site. Protein secondary structure is shown in cartoon and the citrate in stick representation. This figure is taken from Brito et al., 2011.

REFERENCES

Allen, K.N., Dunaway-Mariano, D., 2009. Markers of fitness in a successful enzyme superfamily. Curr. Opin. Struct. Biol. 19, 658–665.

Aravind, L., Galperin, M.Y., Koonin, E.V., 1998. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. Trends Biochem. Sci. 23, 127–129.

Bae-Lee, M.S., Carman, G.M., 1984. Phosphatidylserine synthesis in Saccharomyces cerevisiae. Purification and characterization of membrane-associated phosphatidylserine synthase. J. Biol. Chem. 259, 10857–10862.

Borges, N., Gonçalves, L.G., Rodrigues, M.V., Siopa, F., Ventura, R., Maycock, C., Lamosa, P., Santos, H., 2006. Biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile Archaeoglobus fulgidus in stress adaptation. J. Bacteriol. 188, 8128–8135.

Borges, N., Matsumi, R., Imanaka, T., Atomi, H., Santos, H., 2010. Thermococcus kodakarensis mutants deficient in di-myo-inositol phosphate use aspartate to cope with heat stress. J. Bacteriol. 192, 191–197.

Brito, J.A., Borges, N., Vonrhein, C., Santos, H., Archer, M., 2011. Crystal structure of Archaeoglobus fulgidus CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper)thermophiles. J. Bacteriol. 193, 2177–2185.

Brown, A.D., 1976. Microbial water stress. Bacteriol Rev 40, 803–846.

Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D., Aravind, L., 2006. Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. Journal of Molecular Biology 361, 1003–1034.

Cutler, A.J., Abrams, S.R., Taylor, D.C., Qi, Q., Huang, Y., 2003. Molecular and biochemical characterization of an aminoalcoholphosphotransferase (AAPT1) from Brassica napus©: effects of low temperature and abscisic acid treatments on AAPT

expression in Arabidopsis plants and effects of over-expression of BnAAPT1 in transgenic Arabidopsis. Planta 217, 547–558.

Dai, J., Finci, L., Zhang, C., Lahiri, S., Zhang, G., Peisach, E., Allen, K.N., Dunaway-Mariano, D., 2009. Analysis of the structural determinants underlying discrimination between substrate and solvent in beta-phosphoglucomutase catalysis. Biochemistry 48, 1984–1995.

Dowhan, W., 2013. A retrospective: Use of Escherichia coli as a vehicle to study phospholipid synthesis and function. Biochim. Biophys. Acta 1831, 471–494.

Empadinhas, N., Da Costa, M.S., 2006. Diversity and biosynthesis of compatible solutes in hyper/thermophiles. Int. Microbiol. 9, 199–206.

Fagone, P., Jackowski, S., 2013. Phosphatidylcholine and the CDP-choline cycle. Biochim. Biophys. Acta 1831, 523–532.

Fischl, A.S., Carman, G.M., 1983. Phosphatidylinositol biosynthesis in Saccharomyces cerevisiae: purification and properties of microsome-associated phosphatidylinositol synthase. J. Bacteriol. 154, 304–311.

Fischl, A.S., Homann, M.J., Poole, M.A., Carman, G.M., 1986. Phosphatidylinositol synthase from Saccharomyces cerevisiae. Reconstitution, characterization, and regulation of activity. J. Biol. Chem. 261, 3178–3183.

Gonçalves, L.G., Borges, N., Serra, F., Fernandes, P.L., Dopazo, H., Santos, H., 2012. Evolution of the biosynthesis of di-myo-inositol phosphate, a marker of adaptation to hot marine environments. Environ. Microbiol. 14, 691–701.

Gonçalves, L.G., Huber, R., Da Costa, M.S., Santos, H., 2003. A variant of the hyperthermophile Archaeoglobus fulgidus adapted to grow at high salinity. FEMS Microbiol. Lett. 218, 239–244.

Gopalakrishnan, A.S., Chen, Y.C., Temkin, M., Dowhan, W., 1986. Structure and expression of the gene locus encoding the phosphatidylglycerophosphate synthase of Escherichia coli. J. Biol. Chem. 261, 1329–1338.

Griffin, J.L., Bowler, M.W., Baxter, N.J., Leigh, K.N., Dannatt, H.R.W., Hounslow, A.M., Blackburn, G.M., Webster, C.E., Cliff, M.J., Waltho, J.P., 2012. Near attack conformers dominate β-phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. Proc. Natl. Acad. Sci. U.S.A. 109, 6910–6915.

Henneberry, A.L., McMaster, C.R., 1999. Cloning and expression of a human choline/ethanolaminephosphotransferase: synthesis of phosphatidylcholine and phosphatidylethanolamine. Biochem. J. 339 (Pt 2), 291–298.

Hirabayashi, T., Larson, T.J., Dowhan, W., 1976. Membrane-associated phosphatidylglycerophosphate synthetase from Escherichia coli: purification by substrate affinity chromatography on cytidine 5'-diphospho-1,2-diacyl-sn-glycerol sepharose. Biochemistry 15, 5205–5211.

Hjelmstad, R.H., Bell, R.M., 1990. The sn-1,2-diacylglycerol cholinephosphotransferase of Saccharomyces cerevisiae. Nucleotide sequence, transcriptional mapping, and gene product analysis of the CPT1 gene. J. Biol. Chem. 265, 1755–1764.

Hjelmstad, R.H., Bell, R.M., 1991. sn-1,2-diacylglycerol choline- and ethanolaminephosphotransferases in Saccharomyces cerevisiae. Nucleotide sequence of the EPT1 gene and comparison of the CPT1 and EPT1 gene products. J. Biol. Chem. 266, 5094–5103.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., De Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A.F., Rivoire, C., Sangrador-Vegas, A., Selengut, J.D., Sigrist, C.J.A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P.D., Wu, C.H., Yeats, C., Yong, S.-Y., 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40, D306–312.

Kennedy, E.P., Weiss, S.B., 1956. The function of cytidine coenzymes in the biosynthesis of phospholipides. J. Biol. Chem. 222, 193–214.

Koonin, E.V., Tatusov, R.L., 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. J. Mol. Biol. 244, 125–132.

Lahiri, S.D., Zhang, G., Dunaway-Mariano, D., Allen, K.N., 2002. Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. Biochemistry 41, 8351–8359.

Larson, T.J., Hirabayshi, T., Dowhan, W., 1976. Phosphatidylglycerol biosynthesis in Bacillus licheniformis Resolution of membrane-bound enzymes by affinity chromatography on cytidinediphospho-sn-1,2-diacylglycerol Sepharose. Biochemistry 15, 974–979.

Lykidis, A., 2007. Comparative genomics and evolution of eukaryotic phospholipid biosynthesis. Prog. Lipid Res. 46, 171–199.

Martins, L.O., Santos, H., 1995. Accumulation of Mannosylglycerate and Di-myo-Inositol-Phosphate by Pyrococcus furiosus in Response to Salinity and Temperature. Appl. Environ. Microbiol. 61, 3299–3303.

McMaster, C.R., Morash, S.C., Bell, R.M., 1996. Phospholipid and cation activation of chimaeric choline/ethanolamine phosphotransferases. Biochem. J. 313 (Pt 3), 729–735.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr. D Biol. Crystallogr. 67, 386–394.

Morii, H., Koga, Y., 2003. CDP-2,3-Di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon Methanothermobacter thermautotrophicus. J. Bacteriol. 185, 1181–1189.

Morii, H., Ogawa, M., Fukuda, K., Taniguchi, H., Koga, Y., 2010. A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. J. Biochem. 148, 593–602.

Neves, A.R., Pool, W.A., Castro, R., Mingote, A., Santos, F., Kok, J., Kuipers, O.P., Santos, H., 2006. The alpha-phosphoglucomutase of Lactococcus lactis is unrelated to the alpha-D-phosphohexomutase superfamily and is encoded by the essential gene pgmH. J. Biol. Chem. 281, 36864–36873.

Ohta, A., Waggoner, K., Radominska-Pyrek, A., Dowhan, W., 1981. Cloning of genes involved in membrane lipid synthesis: effects of amplification of phosphatidylglycerophosphate synthase in Escherichia coli. J. Bacteriol. 147, 552–562.

Raetz, C.R., Carman, G.M., Dowhan, W., Jiang, R.T., Waszkuc, W., Loffredo, W., Tsai, M.D., 1987. Phospholipids chiral at phosphorus. Steric

course of the reactions catalyzed by phosphatidylserine synthase from Escherichia coli and yeast. Biochemistry 26, 4022–4027.

Rodionov, D.A., Kurnasov, O.V., Stec, B., Wang, Y., Roberts, M.F., Osterman, A.L., 2007. Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate. Proc. Natl. Acad. Sci. U.S.A. 104, 4279–4284.

Rodrigues, M.V., Borges, N., Henriques, M., Lamosa, P., Ventura, R., Fernandes, C., Empadinhas, N., Maycock, C., Da Costa, M.S., Santos, H., 2007. Bifunctional CTP:inositol-1-phosphate cytidylyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles. J. Bacteriol. 189, 5405–5412.

Santos, H., Lamosa, P., Borges, N., Faria, T., Neves, C., 2007. The physiological role, biosynthesis and mode of action of compatible solutes from (hyper) thermophiles. Physiology and biochemistry of extremophiles, ASM Press, Washington, DC 86–103.

Schollen, E., Pardon, E., Heykants, L., Renard, J., Doggett, N.A., Callen, D.F., Cassiman, J.J., Matthijs, G., 1998. Comparative analysis of the phosphomannomutase genes PMM1, PMM2 and PMM2psi: the sequence variation in the processed pseudogene is a reflection of the mutations found in the functional gene. Hum. Mol. Genet. 7, 157–164.

Shackelford, G.S., Regni, C.A., Beamer, L.J., 2004. Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. Protein Sci. 13, 2130–2138.

Silvaggi, N.R., Zhang, C., Lu, Z., Dai, J., Dunaway-Mariano, D., Allen, K.N., 2006. The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. J. Biol. Chem. 281, 14918–14926.

Sohlenkamp, De Rudder KE, Rohrs, Lopez-Lara, Geiger, 2000. Cloning and characterization of the gene for phosphatidylcholine synthase. J. Biol. Chem. 275, 27500.

Solís-Oviedo, R.L., Martínez-Morales, F., Geiger, O., Sohlenkamp, C., 2012. Functional and topological analysis of phosphatidylcholine synthase from Sinorhizobium meliloti. Biochim. Biophys. Acta 1821, 573–581.

Tomavo, S., Dubremetz, J.F., Schwarz, R.T., 1992. Biosynthesis of glycolipid precursors for glycosylphosphatidylinositol membrane anchors in a Toxoplasma gondii cell-free system. J. Biol. Chem. 267, 21446–21458.

Vance, D.E., 2002. Chapter 8 Phospholipid biosynthesis in eukaryotes, in: New Comprehensive Biochemistry. Elsevier, pp. 205–232.

Williams, J.G., McMaster, C.R., 1998. Scanning alanine mutagenesis of the CDP-alcohol phosphotransferase motif of Saccharomyces cerevisiae cholinephosphotransferase. J. Biol. Chem. 273, 13482–13487.

Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L.W., Allen, K.N., 2005. Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. Biochemistry 44, 9404–9416.

Chapter 2

STRUCTURAL STUDY OF AN ATYPICAL α-PHOSPHOGLUCOMUTASE

The results presented in this chapter were published or submitted:

Przemyslaw Nogly, Rute Castro, Matteo de Rosa, Ana Rute Neves, Helena Santos and Margarida Archer, (2012). Production and crystallization of α-phosphoglucomutase from *Lactococcus lactis*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 68, 1113–1115.

Przemyslaw Nogly, Pedro Matias, Matteo de Rosa, Rute Castro, Helena Santos, Ana Rute Neves, Margarida Archer, (2012). High-resolution structure of an atypical α-phosphoglucomutase related to eukaryotic phosphomannomutases. (Submitted)

NOTE: This work was performed in collaboration with the Lactic acid bacteria & in vivo NMR laboratory (ITQB-UNL), headed by Ana Rute Neves. PN and RC cloned and expressed the protein with help of ARN; PN purified the protein with help of MR; PN crystallized protein with help of MA and MR; PN, PM, MA determined the structure; PN refined the structure with help of MA; PN and MA analyzed the structure and wrote the manuscripts.

2.1 INTRODUCTION

α-Phosphoglucomutase (α-PGM) (E.C. 5.4.2.2) is present in all organisms from bacteria to animals and plants (Whitehouse et al., 1998). This enzyme catalyzes the interconversion of α-glucose 1-phosphate to glucose 6-phosphate, and plays distinct roles among the different organisms. In animals and plants, α-PGM is mainly involved in the synthesis and utilization of storage carbohydrates like glycogen and starch (Fettke et al., 2009; Villar-Palasi and Larner, 1970). In bacteria, α-PGM is involved in sugar utilization and also in the synthesis of UDP-glucose, a sugar donor for the production of glucose-containing polysaccharides. Number of studies associated the virulence of several pathogens to the presence of α-PGM (Paterson et al., 2009; Plant et al., 2006).

Lactococcus lactis is a lactic acid bacterium widely used in starter cultures for the manufacture of fermented dairy products, such as cheese and buttermilk. Its enormous economical value has rendered this typical homofermentative bacterium as one of the most extensively studied members of the lactic acid bacteria. *L. lactis* possesses two distinct PGMs with specificity for α- or β-anomers of phosphoglucose. β-PGM, a member of the haloacid dehalogenase superfamily (HADSF) class I, catalyzes the reversible conversion of β-glucose 1-phosphate to glucose 6-phosphate, and is involved in the degradation pathways of maltose and trehalose (Andersson et al., 2001; Lahiri et al., 2002; Levander et al., 2001). α-PGM of *L. lactis*

(APGM) catalyzes the interconversion of α-glucose 1-phosphate to glucose 6-phosphate, and is essential in the utilization of galactose via the Leloir pathway (Grossiord et al., 1998; Neves et al., 2006). APGM is also implicated in the synthesis of precursors for cell wall polysaccharides and exopolysaccharides (Delcour et al., 1999; Kleerebezem et al., 1999). Therefore, APGM plays an important role in the processes of galactose degradation and formation of precursors for biosynthetic pathways during growth on glucose or other carbohydrates.

In contrast with most bacterial and eukaryotic α-PGMs that utilize α-glucose 1-phosphate and belong to the α-D-phosphohexomutase superfamily (IPR005841), the lactococcal enzyme belongs to the HADSF (Neves et al., 2006). This superfamily comprises more than 3000 proteins, which are mainly involved in phosphoryl transfer reactions (Allen and Dunaway-Mariano, 2004). Members of the HADSF are typically composed of two domains: the core and cap domains. The core domain contains most of the conserved catalytic residues, while the cap domain acts as a lid over the core domain and differentiates according to specific substrates and catalytic reaction (Allen and Dunaway-Mariano, 2009; Lahiri et al., 2004). An example of a strictly conserved amino acid residue is the aspartate that acts as a nucleophile and mediates the transfer of the phosphoryl group (Asp8 in APGM). Members of the HADSF also share the catalytic requirement for a metal cofactor, usually Mg^{2+}. Phylogenetic studies with HADSF proteins suggested a common origin for APGM and the

eukaryotic phosphomannomutases (PMMs; IPR005002) (Neves et al., 2006). However, APGM shows strict specificity for α-glucose 1-phosphate, whereas the eukaryotic phosphomannomutases generally use both α-mannose 1-phosphate and α-glucose 1-phosphate as substrates. Noteworthy, the sequence identity between α- and β-phosphoglucomutase from *L. lactis* is only 10% (belonging to distinct clades of phylogenetic tree; Neves et al., 2006) and results in significant structural differences, namely the topology and position of the cap domain in the sequence. These observations suggest that *L. lactis* APGM is mechanistically closer to the PMMs than to the β-PGMs.

In this work, we report the first crystal structure of APGM determined by single-wavelength anomalous dispersion (SAD) method based on a Pt derivative and refined to 1.5 Å resolution. The structure is compared with other members of the HADSF and specific features of the enzyme with impact on substrate specificity and catalytic mechanism are discussed.

2.2 PRODUCTION AND CRYSTALLIZATION OF α–PHOSPHOGLUCOMUTASE FROM *LACTOCOCCUS LACTIS*

2.2.1 MATERIALS AND METHODS

2.2.1.1 CLONING, EXPRESSION AND PURIFICATION

The nucleotide sequence *pgmH*, encoding APGM was amplified by PCR from *Lactococcus lactis subsp. cremoris*, strain NZ9000, using forward primer 5'CATGCCATGGGCCATCATCATCATCATCAT GACGACGACGACAAGATGAAAAAAATATTAAGT3' and reverse primer 5'GCTCTAGATTAAGCTTCTTCCATCGCAATAATTGCTT TTAGAATAGCTGCAG3'. The gene was cloned into the pNZ8048 vector introducing a hexahistidine (His_6) tag at the N-terminus, followed by an enterokinase cleavage site. The cloned DNA sequence was confirmed to be correct by AGOWA GmbH DNA sequencing service.

Homologous expression was performed in *L. lactis*, strain NZ9000, grown in a 5 liter fermentor with M17 medium (Difco) containing 0.5% glucose, 5 mg l^{-1} chloroamphenicol at 303 K in anaerobic conditions with controlled pH of 6.5. At OD_{600nm} of 0.5 the protein overexpression was induced by adding nisin to a final concentration of 1 μg l^{-1} and subsequently grown for further 2 h. The biomass was harvested by centrifugation at 4000*g*, resuspended in 50 m*M* phosphate buffer pH 8.0, 300 m*M* NaCl, 20 m*M* imidazole and lysed

using a French Press. The cell debris was removed by centrifugation at 30000*g* for 1 h. The soluble extract was loaded onto a 5 ml Ni-NTA Superflow Cartridge (Qiagen). The column was washed with 50 m*M* phosphate buffer pH 8.0, 600 m*M* NaCl, 20 m*M* imidazole until the UV absorbance at 280 nm reached baseline. A gradient elution was used with 50 m*M* phosphate buffer pH 8.0, 300 m*M* NaCl, 500 m*M* imidazole as the elution buffer. The purity of the protein was assessed by SDS-PAGE (Figure 2.1) and its activity was confirmed by a functional assay for α-phosphoglucomutase activity as described previously (Neves et al., 2006) (data not shown). The sample buffer was exchanged to 10 m*M* HEPES pH 7.4 and protein concentrated to about 10 mg ml^{-1} using Vivaspin concentrators with 10 kDa molecular-weight cutoff. The protein sample was divided into 100 µl aliquots, flash frozen in liquid nitrogen and stored at 193 K.

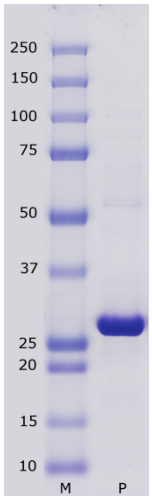


Figure 2.1 SDS-PAGE of purified 30 kDa APGM (construct with His tag) stained with Coomassie blue. Lane M, molecular-weight marker (kDa) from BioRad (Precision Plus Protein Standards); lane P, protein sample (10 µg).

2.2.1.2 CRYSTALLIZATION

Screening of the crystallization conditions was performed with Cartesian Mini-Bee nano-robot at 293 K using the Classics suite and the Classics II suite from Qiagen. The crystallization sitting drops were prepared by mixing 100 nl of protein sample and 100 nl of reservoir solution. Multiple needle-shaped crystals were obtained in the crystallization condition containing 2 *M* ammonium sulfate, 0.1 *M* HEPES pH 7.5 and 2% PEG 400 (Figure 2.2a). Removal of PEG 400 from this condition, led to the formation of multiple plate-like crystals (Figure 2.2b).

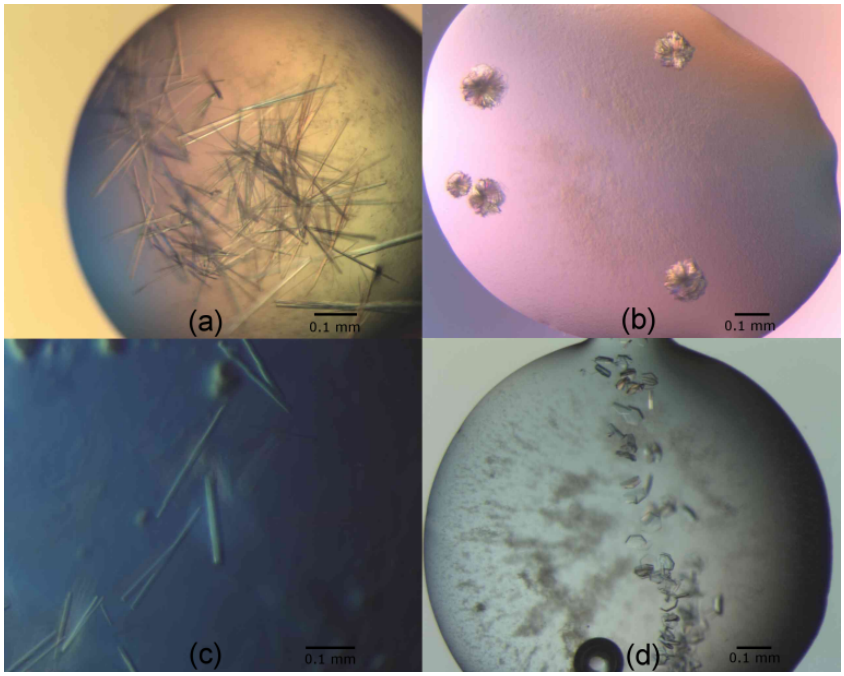


Figure 2.2 Crystals of APGM. (a) Crystals obtained in 2 *M* ammonium sulfate, 0.1 *M* HEPES pH 7.5 and 2% PEG 400. (b) Crystals grown without PEG 400. (c) Crystals obtained using streak seeding into the crystallization solution with 2% PEG 400 from needles; (d) or multiple plates.

Crystal optimization included streak seeding trials from both, needle and plate multiple crystals, into new crystallization solutions comprising 2 *M* ammonium sulfate and 0.1 *M* HEPES pH 7.5. A cat whisker was used to touch the crystal and seeds were immediately transferred in one quick movement across the new drop. Multiple crystals appeared showing similar shapes as the crystals used for seeding (not shown). Furthermore, streak seeding of needles, grown in the presence of PEG 400, into the same crystallization condition yielded single tiny crystals, but not suitable for diffraction experiments (Figure 2.2c).

The crystallization procedure was optimized, so that single hexagonal plate crystals were obtained by cross seeding of multiple plate crystals, grown in the absence of PEG 400 and transferred after 24 h equilibration into crystallization solutions containing 1.9 – 2.1 *M* ammonium sulfate, 0.1 *M* HEPES pH 7.5 and 2% PEG 400 (Figure 2.2d). Hanging drops were prepared by mixing protein and reservoir solutions in ratios 2:1, 1:1 and 1:2 (µl), equilibrated against 500 µl of reservoir at 293 K.

2.2.1.3 DATA COLLECTION AND PROCESSING

Crystals were transferred from the drop with a cryoloop into a new drop containing the crystallization solution supplemented with 20% glycerol and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected on beamline ID 14-2 at the European Synchrotron Radiation Facility (ESRF, Grenoble) at 100 K.

1400 images with an oscillation angle of 0.1° and an exposure time of 1.5 s per frame were recorded with an ADSC Q4 CCD detector at a wavelength of 0.9334 Å. Data were integrated using MOSFLM (Battye et al., 2011) and scaled with SCALA (Evans, 2006). The Matthews coefficient was calculated using the CCP4i package (Collaborative Computational Project, Number 4, 1994; Potterton et al., 2003; Winn et al., 2011).

Table 2.1 Data collection and processing.

Values for the outer resolution shell are given in parentheses.

X-ray source	ESRF, beamline ID 14-2
Wavelength (Å)	0.9334
Temperature (K)	100
Detector	ADSC Q4
Rotation range per image (°)	0.1
Space group	$P3_121/P3_221$
Unit-cell parameters (Å)	a = b = 67.2, c = 210.4
Resolution range (Å)	33.00 - 1.50 (1.58 - 1.50)
Total No. of reflections	744514 (105892)
No. of unique reflections	89350 (12888)
Completeness (%)	100.0 (100.0)
Average multiplicity	8.3 (8.2)
Average I/σ(I)	16.8 (3.1)
R_{meas}	0.089 (0.754)
Overall B factor from Wilson plot (Å^2)	15.9

$R_{meas} = \Sigma_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity, $\langle I(hkl)\rangle$ is the average intensity for multiple measurements and N is the number of times a given reflection has been observed.

2.2.2 RESULTS AND DISCUSSION

The homologous expression of APGM in *L. lactis* was successfully achieved. Purification of APGM by affinity chromatography was sufficient to produce pure and functional protein (Figure 2.1) with a yield of ~ 1.5 mg l^{-1}. The problem of crystal multiplicity observed during the initial screens was overcome by seeding techniques. Single hexagonal crystals were obtained by streak seeding from multiple crystals grown in ammonium sulfate in the absence of PEG 400 into the same crystallization solution supplemented with 2% PEG 400. Crystals grew up to 150 µm within one week after the streak seeding (Figure 2.2d). Later on, these new single crystals obtained from seeding were used as the initial source of seeds. Crystals diffracted to 1.5 Å resolution and belonged to the trigonal space group $P3_{1,2}21$ with unit cell parameters a=b= 67.17 Å and c= 210.39 Å. The calculated Matthews coefficient (Kantardjieff and Rupp, 2003; Matthews, 1968) is 2.3 $Å^3$/Da with a solvent content around 46% indicating the presence of two molecules in the asymmetric unit. The relevant parameters for data collection and processing are summarized in Table 2.1.

A BLAST search revealed highest identity of APGM sequence with phosphomannomutases (around 22-24%). The best three hits, for which the three-dimensional structure are available, were used as search models for molecular replacement (human phosphomannomutase 2, PDB code 2amy, to be published; human

phosphomannomutase 1, PDB code 2fue, (Silvaggi et al., 2006); and *Leishmania* phosphomannomutase, PDB code 2i54, (Kedzierski et al., 2006)). No solution was found in the various molecular replacement trials with Molrep (Vagin and Teplyakov, 1997) and Phaser (McCoy et al., 2007). Therefore, the phase problem still needed to be addressed by experimental phasing, which is described in Section 2.3.1.4.

2.3 HIGH-RESOLUTION STRUCTURE OF AN ATYPICAL α–PHOSPHOGLUCOMUTASE

2.3.1 MATERIALS AND METHODS

2.3.1.1 CRYSTALLIZATION

The crystallization of APGM was performed as described in section 2.2.1.2 (Nogly et al., 2012). Native crystals were soaked at room temperature in crystallization solutions supplemented with a heavy-atom compound. The soaking times were 20 h for K_2**Pt**Cl_4, K_2**Pt**$(NO_2)_4$, $C_5H_{11}O_2$**Pb**, C_9H_9**Hg**NaO_2S and K**Au**Cl_4, and 1 month for **Hg**$(CH_3COO)_2$ and **Pt**$(NH_3)_2Cl_2$. The final concentration of these compounds ranged from 2.5 to 15 mM.

2.3.1.2 GEL FILTRATION

Size-exclusion chromatography was carried out to estimate the molecular mass of APGM in solution, using a Superdex 200 16/60 gel filtration column equilibrated with 50 mM HEPES pH 7.4 and 150 mM NaCl. The calibration curve was prepared using as standards cytochrome *c*, carbonic anhydrase, albumin (BSA), alcohol dehydrogenase and β-amylase.

2.3.1.3 DATA COLLECTION AND PROCESSING

The data set from the crystal derivatized with K_2PtCl_4 was collected to 2.4 Å resolution at the I04 beamline of the Diamond Light Source

(DLS, Didcot), at 100 K, and the diffraction data were integrated and scaled with XDS (Kabsch, 2010). The other putative derivative data sets (data not shown) were measured at the ESRF, DLS or SLS (Swiss Light Source, Villigen) and processed with XDS.

2.3.1.4 PHASING AND CRYSTALLOGRAPHIC REFINEMENT

HKL2MAP graphic interface (Pape and Schneider, 2004) and SHELXD (Schneider and Sheldrick, 2002) were used to search all the putative derivatives for heavy atom sites. The only clear solution was observed for the K_2PtCl_4 derivative, yielding 23 positive hits out of 100 trials, the best solution showing correlation coefficients CC_{all} = 37.6% and CC_{weak} = 21.6%. A total of 5 probable Pt sites were found with occupancies higher than 0.3 in space group $P3_121$. However, SHELXE (Sheldrick, 2002) did not produce an interpretable electron density map. Therefore, the initial set of 5 sites was input to SHARP/autoSHARP (Bricogne et al., 2003; Vonrhein et al., 2007) for further refinement and phasing, with phasing power (anomalous) of 1.622 and figure of merit (acentric) 0.343. After density modification with Solomon (Abrahams and Leslie, 1996) an interpretable electron density map was obtained at 2.4 Å resolution. The phase extension to include the higher resolution native data set was not successful, very likely due to the non-isomorphism between the Pt derivative and the native crystals. Nevertheless, a partial backbone model of the protein was built with Buccaneer (Cowtan, 2006). Based on this incomplete model, it was possible to identify the non-crystallographic symmetry

(NCS) between the two molecules in the asymmetric unit. The NCS symmetry operator was used for density modification in DM (Cowtan, 1994) which further improved the quality of the calculated electron density maps. Cycles of automated model building with Buccaneer, ARP/wARP (Langer et al., 2008) alternating with manual model building using Coot (Emsley et al., 2010), yielded a crystallographic model about 80% complete. Molecular replacement was then successfully carried out with Phaser (McCoy et al., 2007) using this partial model as a search template against the higher resolution native data set. The structure was completed and refined to 1.5 Å resolution using Coot and Refmac5 (Murshudov et al., 1999). In the final refinement cycles, the model was refined using TLS groups as generated with TLSMD server (Painter and Merritt, 2006) and with hydrogen atoms added in the ideal positions with MolProbity server (Chen et al., 2010) used as riding atoms.

Table 2.2 Data collection and processing of Pt derivative and structure refinement.

Data collection and processing of Pt derivative data	
X-ray source	DLS, beamline I04
Wavelength (Å)	0.9722
Temperature (K)	100
Detector	ADSC Quantum Q315r
Rotation range per image (°)	0.4
Space group	$P3_121$
Unit-cell parameters (Å)	a = b = 67.4, c = 212.6
Resolution range (Å)	50.00 - 2.40 (2.54 – 2.40)
Total No. of reflections	129424 (17888)
No. of unique reflections	41740 (6489)
Completeness (%)	98.5 (95.1)
Average multiplicity	3.1 (2.8)
Average I/σ(I)	25.76 (10.91)
R_{meas}	0.034 (0.109)
Overall B factor from Wilson plot (Å^2)	42.6
Statistics on refinement of native data	
Resolution (Å)	30.0-1.50
No. reflections	89298
R_{work} / R_{free}	0.155/ 0.178
R.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.752

Statistics are given treating Bijvoet mates as separate observations.

Values for the higher resolution shell are given in parentheses.

$R_{meas} = \Sigma_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity, $\langle I(hkl)\rangle$ is the average intensity for multiple measurements and N is the number of times a given reflection has been observed.

2.3.2 RESULTS AND DISCUSSION

2.3.2.1 MODEL QUALITY AND APGM STRUCTURE

The APGM crystal soaked with K_2PtCl_4 diffracted to 2.4 Å and belonged to space group $P3_121$ with unit cell dimensions of a = b = 67.4 Å and c = 212.6 Å. Two molecules are present in the asymmetric unit corresponding to a solvent content around 47.4%. The structure of APGM was solved by single wavelength anomalous scattering (SAD) using the K_2PtCl_4 derivative. Relevant parameters for data collection and processing of the derivative data are shown in Table 2.2. The final APGM model was refined to 1.5 Å resolution with R_{factor} of 15.5% and R_{free} of 17.8% (Brünger, 1992). The crystallographic model showed good geometry (rmsd bond length 0.0159, rmsd bond angle 1.7482) and stereochemistry (PROCHECK, Laskowski et al., 1993). All amino acid residues lie within the allowed regions of the Ramachandran plot (RAMPAGE, Lovell et al., 2003).

The electron density maps are very well defined for the protein backbone, except for the region comprising Ile9 and Asp10 in molecule B. The side-chains of 12 surface residues are not fully visible in the final 2|Fo|-|Fc| electron density map contoured at 1σ, likely due to their mobility, and 27 side-chains were modeled with alternate conformations, including the catalytic residue Asp8. A blob of residual electron density was observed near the active site of the enzyme (cap domain). Based on its tetrahedral shape and nearby residues, a sulfate ion was fitted into the blob and further refined. Six glycerol molecules were added to the model, one of which is close to

the sulfate ion in chain B. The final APGM model comprises 504 amino acid residues (252 per chain), 606 water molecules, 2 sulfate ions and 6 glycerol molecules.

Each molecule comprises a core domain (residues 1-88 and 189-252), typical of HADSF, and a small cap domain (residues 93-183) as illustrated in Figure 2.3.

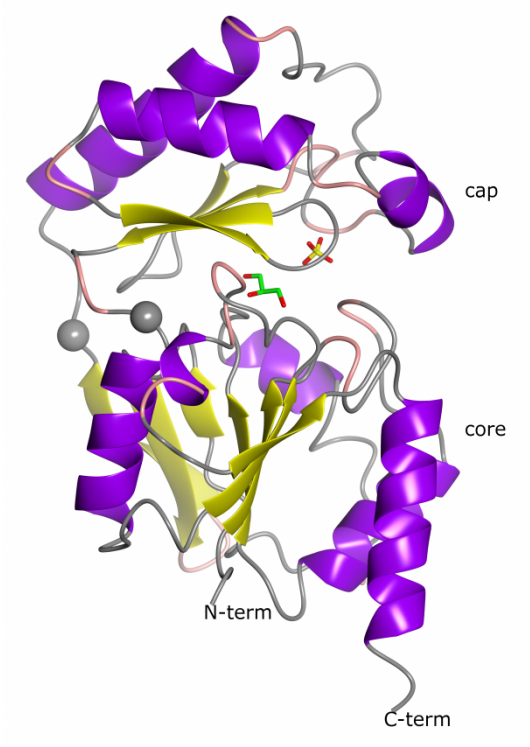


Figure 2.3 APGM ribbon representation (molecule B), colored according to secondary structure elements assigned with DSSP (Kabsch and Sander, 1983) and hinges by HingeProt (yellow β-strands, magenta α-helices, pink turns, sphere hinges). Glycerol and sulfate located at the interface of the core and cap domains are drawn in stick mode. Atom color codes used in this chapter, unless stated otherwise: C - green, O - red and S – yellow. Figures were drawn with CCP4MG (McNicholas et al., 2011).

The core domain consists of a seven-stranded β-sheet (5, 4, 3, 2, 1, 10 and 11) flanked by five α-helices (6, 7, 8, 1 and 2; refer to Figure 2.5 for numbering). The cap domain is constituted by a four-stranded antiparallel β-sheet (6, 7, 9 and 8) with three α-helices located on one face (3, 5 and 4). The cap and core domains are connected by two linker loops (89-92 and 184-188) that may act as a hinge joint. Indeed, HingeProt (Emekli et al., 2008) predicts a more mobile region around residues 88 and 188 (Figure 2.3).

The APGM active site is located at the interface between the core and cap domains. This interface comprises several positively charged residues, such as Lys17 and Lys44 in the core domain, and Arg128, Lys140 and Lys146 in the cap domain, which are proposed to keep the enzyme in an open conformation. Upon binding of the negatively charged substrate (i.e. α-glucose 1-phosphate) in cap domain, the electrostatic repulsion between the cap and core domains is shielded, favoring the closure of the cap domain and proper positioning of the substrate with respect to catalytic residues in the core domain. A closed cap domain also provides a solvent-excluded environment. The core domain contains the conserved catalytic residues of the HADSF (see more details below), while the cap domain has more variable size and structure, and thus is proposed to be involved in substrate recognition. The topology and location of the cap domain allows the classification of APGM within the type IIb subfamily of the HADSF (Allen and Dunaway-Mariano, 2004).

2.3.2.2 HOMOLOGOUS STRUCTURES

The overall structure of APGM is similar to eukaryotic phosphomannomutases (PMMs), despite their low sequence identity (up to 24%). The four best hits in DALI were (PDB entries, Z-scores and rmsd for C^{α} superposition are indicated for each hit): *Leishmania mexicana* PMM (2i55, 23.7, 2.9 Å), *Trypanosoma brucei* PMM (3f9r, 24.2; 3.0 Å), human PMM1 (2fue, 20.1, 4.1 Å) and human PMM2 (2amy, 17.4, 5.3 Å (Kedzierski et al., 2006; Silvaggi et al., 2006). Superposition of APGM with these four crystal structures shows that the cap domain is more tightly closed in APGM than in PMMs (Figure 2.4).

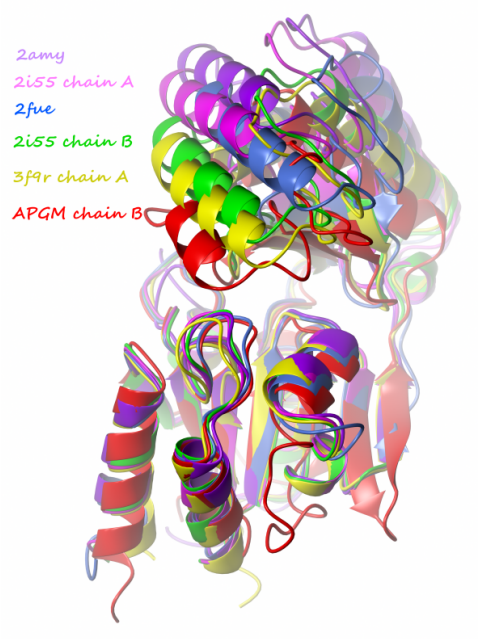


Figure 2.4 Structural superposition of APGM and PMMs aligned by the core domains (with SwissPDBViewer, Guex and Peitsch, 1997) showing the different conformations of the cap domain relative to the core domain. APGM (depicted in red) being the most closed one. The PDB code identifies the PMM structures: 2i55 - *Leishmania* Mexicana PMM, 2fue - human PMM1, 2amy - human PMM2 and 3f9r - *Trypanosoma brucei* PMM.

In contrast, the structure of human PMM2 (2amy) without any ligand bound at the domains interface, has the cap domain furthest away from the core domain. The structure of leishmanial PMM (2i55) illustrates the cap closure upon ligand binding: polypeptide chain A has no ligand at the active site, whereas chain B has a ligand bound and thus is in a more closed conformation. This ligand, glucose 1,6-bisphosphate, is extensively interacting with both domains. Remarkably, the structure of human PMM1 bound to mannose 1-phosphate (2fue) was captured in a cap open state, with the substrate interacting mainly with residues from the cap domain. This structure was obtained by soaking the ligand into pre-formed crystals, which was suggested to restrict the movements of the cap domain within the crystal lattice (Silvaggi et al., 2006).

Sulfate binding at the active site, as in APGM and in 3f9r, rather than larger ligands, as in 2i55 and 2fue, is likely to allow a more closed conformation of the cap domain. However, the structure of *T. brucei* PMM with sulfate displays a more open conformation than APGM. The positive patch at the cap-core interface of *Trypanosoma* PMM comprises Arg19 and Lys50 in the core domain, and Arg122, Arg133 and Arg140 in the cap domain, and is highly conserved among PMMs. However, in APGM, the Lys140 side-chain (corresponding to Arg133 of 3f9r) is oriented towards the solvent, which could mitigate the electrostatic repulsion between the positively charged residues at the APGM domains interface, thus allowing a more tightly closed cap conformation.

APGM shows several insertion regions with regard to PMMs (Figure 2.5), located mainly at the molecular surface of the core domain, namely between α2 – β3, β4 – β5, α3 – β6 (cap domain) and α6 – β10. These insertions are not expected to cause major structural differences in APGM as compared to PMMs, except for the one at the cap domain, as discussed in the next section.

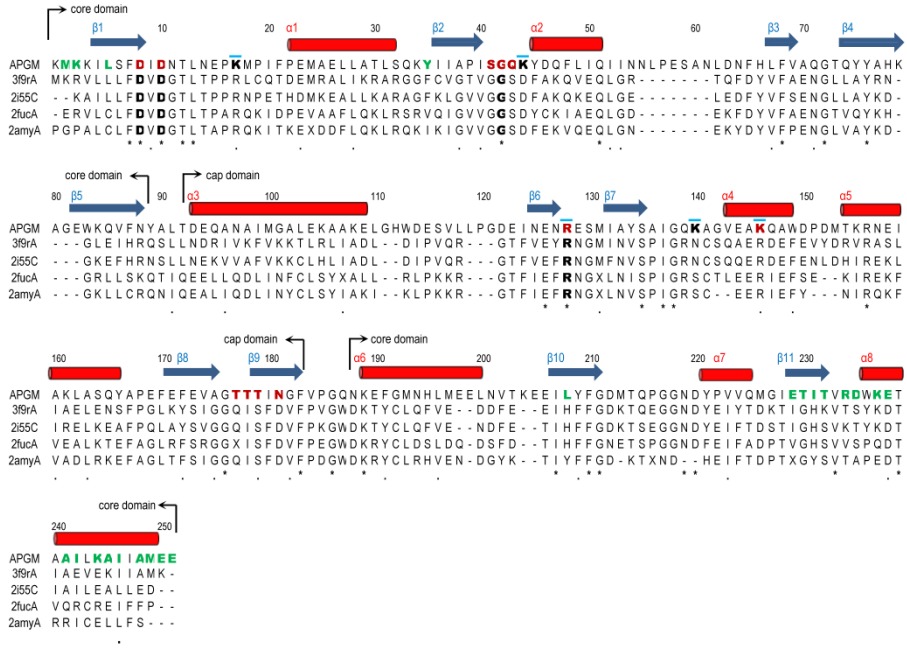


Figure 2.5 Amino acid sequence alignment based on structural superimposition of APGM with PMMs performed with the Dali server. Black arrows indicate domain boundaries. Relevant amino acid residues are displayed in bold; residue color code: red – involved in substrate interaction, blue overline - positive charged residues at cap-core interface, green – forming dimer interface. Below the sequence alignment, conserved residues are indicated by an asterisk (*) and similar residues by a dot (.). Blue arrows represent β-strands and red tubes α-helices. PDB codes refer to structures of: 3f9r – *Trypanosoma brucei* PMM, 2i55 – *Leishmania Mexicana* PMM, 2fuc – human PMM1, 2amy – human PMM2.

2.3.2.3 QUATERNARY STRUCTURE

The crystal asymmetric unit of APGM contains two molecules; the rmsd upon C^{α} superposition of both molecules (chains A and B) is 0.35 Å, indicating very similar structures. The molecular mass of APGM estimated by size-exclusion chromatography was 55 kDa, indicating a dimeric form in solution. The dimeric assembly proposed by PISA (Krissinel and Henrick, 2007) corresponds to a crystal contact of APGM structure (Figure 2.6). The achieved score (CSS) suggests an essential role of the interface in complex formation.

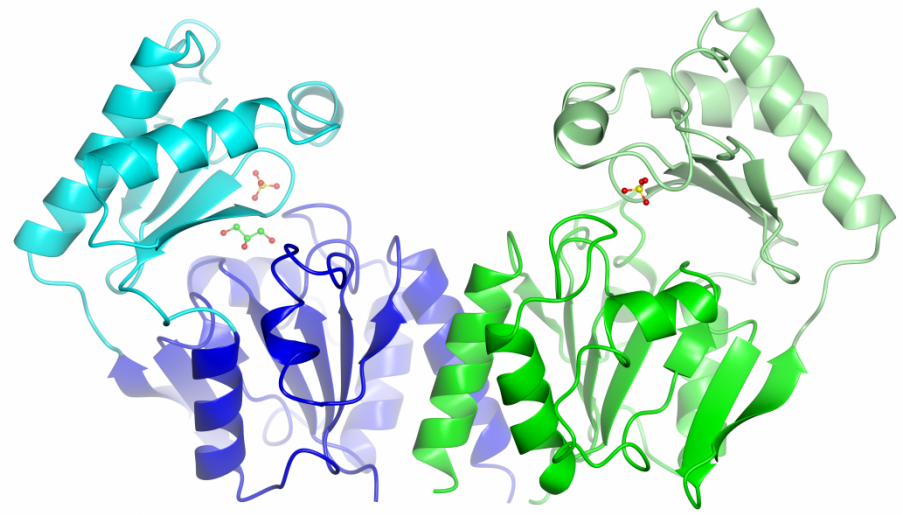


Figure 2.6 APGM dimer identified by PISA. Molecule B is colored in blue and molecule A in green (dark color for the core and light for the cap domain). Sulfate and glycerol are located at the substrate binding site.

The occluded area at this dimer interface is approximately 1000 $Å^2$ (around 5% of the total area) and is formed by hydrophobic amino acid residues located mainly at the C-terminal helix α8, and

hydrophilic residues sited on the stretch between β11 and α8 (core domain). It also includes waters mediating H-bonds between the two molecules. This dimeric arrangement should represent the biological unit of APGM and may be relevant for the enzyme function, as it brings the active sites within the dimer into proximity. It is worth noting that in APGM, helix α8 is more shifted away from the β-sheet of the same monomer towards α8 and β11 of the other monomer than in PMMs, and thus forming a hydrophobic groove accommodating residues sited on α_8 of the other molecule within the dimer.

The dimeric assembly proposed for *L. mexicana* PMM (Figure 2.7) shows a different interface formed exclusively by amino acid residues from the cap domains (Kedzierski et al., 2006). The dimer formation involves a helix-helix interaction between the first helices (α3) and an antiparallel β-sheet interaction between the first strands (β6) of the two cap domains, which extends the solvent exposed face of the β-sheet, placing the two active sites adjacent to each other. This interface is also conserved in human PMMs (Silvaggi et al., 2006) and likely in *T. brucei* PMM (as predicted by PISA server).

Structure based sequence alignment of APGM with these four homologous PMMs shows a 5-residue-insertion in this region of the cap domain, between helix α3 and strand β6 (Figure 2.5). Structural superposition reveals that APGM adopts a different conformation for the loop comprising Leu110 to Glu123, including a protrusion (Glu115-Glu123) facing outwards, which could indicate that APGM

may not form a similar dimeric assembly as the aforementioned PMMs structures (Figure 2.7).

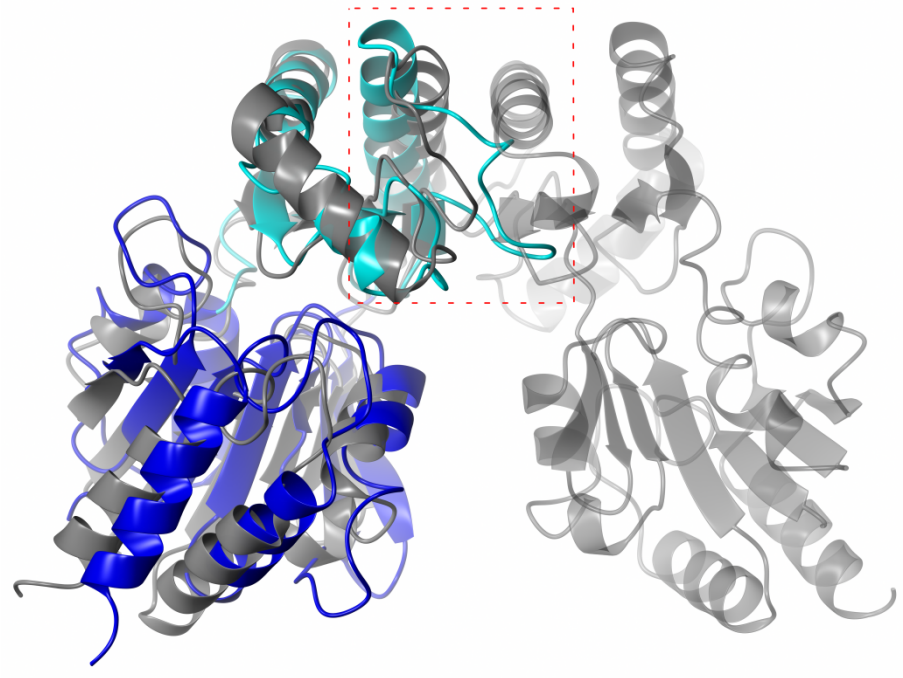


Figure 2.7 Dimer of leishmanial PMM (grey) illustrating the conserved dimeric interface of eukaryotic PMMs. Superimposition with one molecule of APGM (core - dark blue, cap - cyan) by cap domain highlights the longer loop in APGM.

2.3.2.4 ACTIVE SITE AND CATALYTIC MECHANISM

APGM catalyzes the interconversion of α-glucose 1-phosphate to glucose 6-phosphate via a phophorylated enzyme intermediate, as proposed for other phosphryl transfer enzymes of HADSF (Dai et al., 2006; Kedzierski et al., 2006). The enzyme has to be first activated by transfer of a phosphate from the α-glucose 1,6-bisphosphate to the enzyme aspartyl nucleophile (Asp8 in APGM) producing an aspartyl-

phosphate. The substrate (α-glucose 1-phosphate) binds, forming a covalent bond between the substrate C6 hydroxyl oxygen and the aspartyl-phosphate phosphorous. A α-glucose-1,6-bisphosphate intermediate is generated, which then needs to be released from the enzyme and rebound in a reoriented position so that the 1-phosphate is transferred to the aspartyl nucleophile, replenishing the activated form of the enzyme and forming the final product (glucose 6-phosphate). Release of the phosphate from the bisphospho-carbohydrate intermediate to the aspartyl is facilitated by another acid/base group (Asp10). The catalytic nucleophile (Asp8) usually coordinates a magnesium ion along with other residues and water molecules. No metal ion is observed in the APGM crystal structure, and no Mg^{2+} ions were present in the crystallization buffer.

Although the relevance of the APGM dimer to its catalytic activity has not yet been demonstrated, in this arrangement the two active sites are placed face-to-face, with the sulfate ions ~38 Å apart. Interestingly, an extended positively charged patch is observed in a groove between the active sites, comprising two arginines (Arg234A/B) and two lysines (Lys237A/B) residues. This structural configuration can contribute to an efficient turnover, if the bisphosphorylated sugar intermediate can be released from the one active site and immediately captured (in the inverted orientation) by the nearby active site of the second monomer in the dimer. Moreover, the positive charge between the active sites may prolong the presence of the intermediate within the dimer instead of being released to the

solution. A functional role was also proposed for the dimeric form of human and leishmanial phosphomannomutases (Kedzierski et al., 2006; Silvaggi et al., 2006).

The unsuccessful attempts to obtain the structure of APGM in complex with substrate analogues, either by crystal soaking or co-crystallization may be due to the presence of sulfate at the APGM active site (Lu et al., 2005). Nevertheless, important information on substrate binding can be derived from the homologous PMM structures, in particular those of human PMM1 in complex with α-mannose-1-phosphate and leishmanial PMM with β-glucose-1,6-bisphosphate. In fact, the sulfate bound at the active site of APGM molecules A and B overlays quite well with the phosphoryl group (bound to C1) of glucose-1,6-bisphosphate bound to the *L. mexicana* PMM structure (Figure 2.8). Moreover, if the cap domains of APGM and human PMM1 structures are superposed, the sulfate ion in APGM also matches the position of the phosphoryl group of α-mannose-1-phosphate. In the APGM structure, the sulfate is H-bonded to main-chain amide and side-chain hydroxyl groups of Thr177, Thr178, Thr179, Lys146 and a few water molecules. It is also close to Gln43 from the core domain (~4.1 and 3.1 Å in molecules A and B, respectively). In all four eukaryotic PMM structures from the DALI survey, this distal sulfate/substrate phosphoryl binding pocket (cap domain) is composed mainly by two arginines and one serine (Arg143, Arg150 and Ser188, human PMM1 numbering).

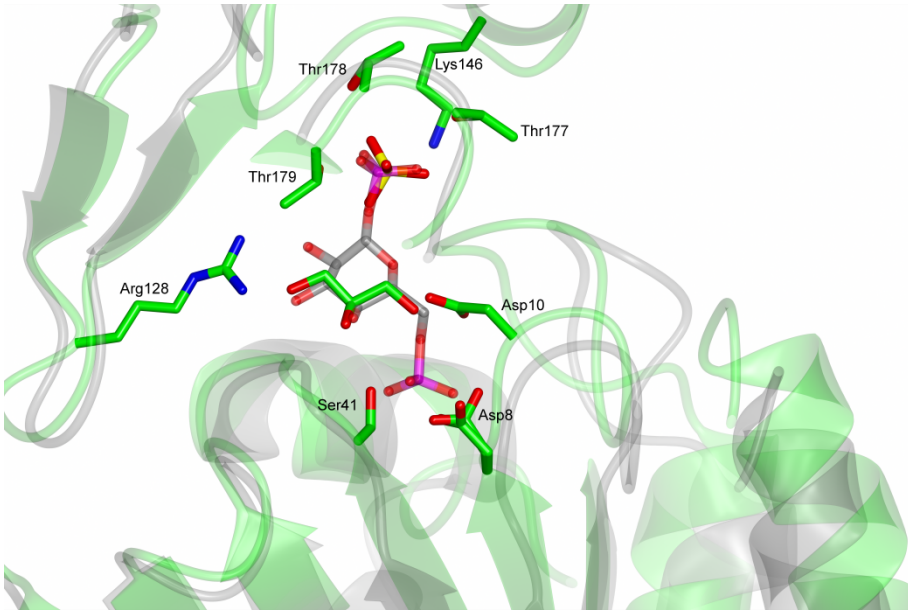


Figure 2.8 Superposition of APGM and leishmanial PMM structures, highlighting the alignment of glycerol and sulfate (APGM; C - green; alternate conformation in Asp8) with glucose 1,6-bisphosphate (leishmanial PMM; C – grey; P - magenta).

In addition, the glycerol found near the sulfate ion (molecule B, closest distance is 4.1 Å) also aligns very well with part of the pyranose ring of β-glucose-1,6-biphosphate bound to the leishmanial PMM structure. In APGM, the two hydroxyl groups of glycerol that mimic part of the sugar moiety establish hydrogen bonds with the side-chains of Arg128 and Asn181 (corresponding to Arg122 and Asp180 in leishmanial PMM), main-chain amide group peptide hydrogen of Gly42 and a few water molecules. It is worth noting that in the APGM structure, the side-chain of Asp10 is H-bonded to one glycerol hydroxyl (corresponding to C6 hydroxyl of the PMM ligand) in molecule B, while in molecule A, in the absence of glycerol, it is 2.8 Å shifted towards the interior of the active site pocket. On the

contrary, in PMMs, the corresponding Asp side-chain is further away from the ligand (~7 Å from the C6 atom of glucose in leishmanial PMM), except for the human PMM1 where the acid/base Asp is in a conformation corresponding to molecule B of APGM. The mobility of Asp10, which plays an important role in the catalytic mechanism, is clearly evident by the different positions adopted by this acid/base residue in APGM and PMMs. Also, the poorly defined electron density around the backbones of Ile9 and Asp10 in molecule B, suggests flexibility and possibly alternate conformations.

Moreover, we have fitted α-glucose-1,6-biphosphate into APGM, based on the previous superposition with leishmanial PMM, and a model with reasonable stereochemistry was obtained except for Ser41, which lies too close (1.3 Å) to the phosphoryl group (Figure 2.8). This residue is replaced by a conserved glycine in PMMs. A shift of the molecule towards Asp8 and a slight rotation around the phosphodiester bond followed by geometry idealization with refmac5 (Murshudov et al., 1997), provided a good fit of the ligand into the binding pocket (Figure 2.9).

A relevant point to discuss is the strict specificity of APGM towards α-glucose 1-phosphate, in contrast to PMMs that can use both α-glucose 1-phosphate and α-mannose 1-phosphate as substrates. The only difference between these two sugars is the position of the hydroxyl group on C2 of the pyranose ring (equatorial in α-glucose 1-phosphate and axial in α-mannose 1-phosphate).

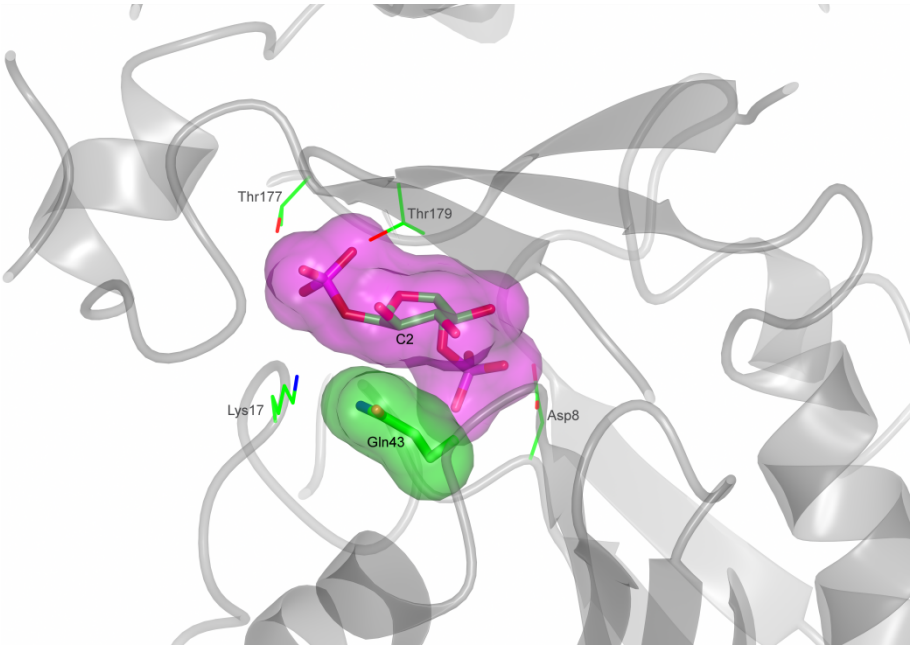


Figure 2.9 Active site of APGM fitted with glucose 1,6-bisphosphate (manual fitting followed by structure idealization). Van der Waals surfaces of Gln43 side-chain and phosphosugar ligand are colored in green and pink, respectively, highlighting the steric restrictions imposed by Gln43 below the C2 atom of the sugar ring.

Structural superimposition of APGM with leishmanial PMM shows that the side-chain of Gln43 in APGM lies in close proximity to C2 of the pyranose ring of α-glucose-1,6-bisphosphate (Gln43 $N^{\varepsilon 2}$ is only 2.3 Å away from C2). This residue restricts the space under the sugar ring, which may be needed to accommodate the mannose 1-phosphate (Figure 2.9). If the hydroxyl group of C2 is in the axial position, as in α-mannose-1,6-bisphosphate, it would clash with the side-chain of Gln43. This steric hindrance in the presence of mannose 1-phosphate may prevent the full closure of the cap over the core domain and restrict the proper positioning of the phosphomannose to react with the catalytic residues in the core domain. This Gln43 is

replaced by a strictly conserved serine in PMMs, and no steric hindrance is observed. This amino acid substitution is likely to be responsible for the strict specificity of APGM towards α-glucose 1-phosphate.

In addition, the crystal structure of human PMM1 with α-mannose 1-phosphate was proposed to represent the first encounter state of the ligand with the enzyme (Silvaggi et al., 2006). The cap domain is found in an open conformation, and the residues engaged in the interactions with the ligand are mainly from the cap domain. In this structure, the hydroxyl group at C2 of the pyranose ring is H-bonded to Arg132, Asn137 and Asp190. The corresponding residues in APGM are Arg128, Ala133 and Asn181. The substitution of Asn to Ala affects the enzyme-substrate interactions, as no hydrogen bond can be established by Ala in APGM. Because the cap domain accounts for first step in substrate recognition and binding, the weaker interactions with α-mannose 1-phosphate in APGM may also contribute for its specificity towards α-glucose 1-phosphate.

CONCLUSIONS

Herein we presented the first high resolution structure of *Lactococcus lactis* α-phosphoglucomutase (APGM) with sulfate and glycerol bound at the active site. The enzyme cap domain is found in a tightly closed conformation. APGM is found to be a dimer in solution as well as in the crystal structure, which might be relevant for the enzyme efficiency; however the dimer interface differs from that found in eukaryotic PMMs. Despite low sequence identity, APGM shares structural homology with phosphomannomutases, although it can only utilize α-glucose 1-phosphate, but not its stereo-isomer α-mannose 1-phosphate, like PMMs. Structural superposition of APGM and PMMs highlights their similarity and differences. A relevant difference seems to be the replacement of conserved Gly and Ser residues in PMMs by Ser41 and Gln43, respectively, in APGM, which adds steric constraints at the APGM active site.

This study provides a framework towards engineering APGM to further understand its catalytic mechanism and strict substrate selectivity. The present lack of success in obtaining APGM crystals with ligands can be attributed to the crystallization conditions used, therefore new crystallization conditions facilitating complex formation of the APGM with ligands will have to be found.

ACKNOWLEDGMENTS

The authors would like to thank to Carlos Frazão and Alekos Athanasiadis for data collection, Luis Fonseca for technical help with the fermentor. This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through grants PTDC/BIA-PRO/103718/2008 and PEst-OE/EQB/LA0004/2011 and by an EU Grant (FP7/2007-2013 No. 211800). We acknowledge Diamond, ESRF and SLS synchrotron for financial support for data collection.

REFERENCES

Abrahams, J.P., Leslie, A.G., 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D Biol. Crystallogr. 52, 30–42.

Allen, K.N., Dunaway-Mariano, D., 2004. Phosphoryl group transfer: evolution of a catalytic scaffold. Trends Biochem. Sci. 29, 495–503.

Allen, K.N., Dunaway-Mariano, D., 2009. Markers of fitness in a successful enzyme superfamily. Curr. Opin. Struct. Biol. 19, 658–665.

Andersson, U., Levander, F., Rådström, P., 2001. Trehalose-6-phosphate phosphorylase is part of a novel metabolic pathway for trehalose utilization in Lactococcus lactis. J. Biol. Chem. 276, 42707–42713.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Cryst. D 67, 271–281.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W., 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. 59, 2023–2030.

Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355, 472–475.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21.

Collaborative Computational Project, Number 4, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763.

Cowtan, K., 1994. "dm": An automated procedure for phase improvement by density modification. CCP4 and ESF-EACBM Newsletter on Protein Crystallography 34–38.

Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. 62, 1002–1011.

Dai, J., Wang, L., Allen, K.N., Radstrom, P., Dunaway-Mariano, D., 2006. Conformational cycling in beta-phosphoglucomutase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate. Biochemistry 45, 7818–7824.

Delcour, J., Ferain, T., Deghorain, M., Palumbo, E., Hols, P., 1999. The biosynthesis and functionality of the cell-wall of lactic acid bacteria. Antonie Van Leeuwenhoek 76, 159–184.

Emekli, U., Schneidman-Duhovny, D., Wolfson, H.J., Nussinov, R., Haliloglu, T., 2008. HingeProt: automated prediction of hinges in protein structures. Proteins 70, 1219–1227.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501.

Evans, P., 2006. Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. 62, 72–82.

Fettke, J., Hejazi, M., Smirnova, J., Höchel, E., Stage, M., Steup, M., 2009. Eukaryotic starch degradation: integration of plastidial and cytosolic pathways. J. Exp. Bot. 60, 2907–2922.

Grossiord, B., Vaughan, E., Luesink, E., De Vos, W., 1998. Genetics of galactose utilisation via the Leloir pathway in lactic acid bacteria. Lait 78, 77–84.

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Kabsch, W., 2010. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Kantardjieff, K.A., Rupp, B., 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 12, 1865–1871.

Kedzierski, L., Malby, R.L., Smith, B.J., Perugini, M.A., Hodder, A.N., Ilg, T., Colman, P.M., Handman, E., 2006. Structure of Leishmania mexicana phosphomannomutase highlights similarities with human isoforms. J. Mol. Biol. 363, 215–227.

Kleerebezem, M., Van Kranenburg, R., Tuinier, R., Boels, I.C., Zoon, P., Looijesteijn, E., Hugenholtz, J., De Vos, W.M., 1999. Exopolysaccharides produced by Lactococcus lactis: from genetic engineering to improved rheological properties? Antonie Van Leeuwenhoek 76, 357–365.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Lahiri, S.D., Zhang, G., Dai, J., Dunaway-Mariano, D., Allen, K.N., 2004. Analysis of the substrate specificity loop of the HAD superfamily cap domain. Biochemistry 43, 2812–2820.

Lahiri, S.D., Zhang, G., Dunaway-Mariano, D., Allen, K.N., 2002. Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. Biochemistry 41, 8351–8359.

Langer, G., Cohen, S.X., Lamzin, V.S., Perrakis, A., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 3, 1171–1179.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26, 283–291.

Levander, F., Andersson, U., Rådström, P., 2001. Physiological role of beta-phosphoglucomutase in Lactococcus lactis. Appl. Environ. Microbiol. 67, 4546–4553.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, De Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50, 437–450.

Lu, Z., Dunaway-Mariano, D., Allen, K.N., 2005. HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. Biochemistry 44, 8684–8696.

Matthews, B.W., 1968. Solvent content of protein crystals. J. Mol. Biol. 33, 491–497.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr. D Biol. Crystallogr. 67, 386–394.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. 55, 247–255.

Neves, A.R., Pool, W.A., Castro, R., Mingote, A., Santos, F., Kok, J., Kuipers, O.P., Santos, H., 2006. The alpha-phosphoglucomutase of Lactococcus lactis is unrelated to the alpha-D-phosphohexomutase superfamily and is encoded by the essential gene pgmH. J. Biol. Chem. 281, 36864–36873.

Nogly, P., Castro, R., De Rosa, M., Neves, A.R., Santos, H., Archer, M., 2012. Production and crystallization of α-phosphoglucomutase from Lactococcus lactis. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 68, 1113–1115.

Painter, J., Merritt, E.A., 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. 62, 439–450.

Pape, T., Schneider, T., 2004. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs RID B-7442-2011. J. Appl. Crystallogr. 37, 843–844.

Paterson, G.K., Cone, D.B., Peters, S.E., Maskell, D.J., 2009. The enzyme phosphoglucomutase (Pgm) is required by Salmonella enterica serovar Typhimurium for O-antigen production, resistance to antimicrobial peptides and in vivo fitness. Microbiology (Reading, Engl.) 155, 3403–3410.

Plant, L., Sundqvist, J., Zughaier, S., Lövkvist, L., Stephens, D.S., Jonsson, A.-B., 2006. Lipooligosaccharide structure contributes to multiple steps in the virulence of Neisseria meningitidis. Infect. Immun. 74, 1360–1367.

Potterton, E., Briggs, P., Turkenburg, M., Dodson, E., 2003. A graphical user interface to the CCP4 program suite. Acta Crystallogr. D Biol. Crystallogr. 59, 1131–1137.

Schneider, T.R., Sheldrick, G.M., 2002. Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. 58, 1772–1779.

Sheldrick, G.M., 2002. Macromolecular phasing with SHELXE. Zeitschrift für Kristallographie 217, 644–650.

Silvaggi, N.R., Zhang, C., Lu, Z., Dai, J., Dunaway-Mariano, D., Allen, K.N., 2006. The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. J. Biol. Chem. 281, 14918–14926.

Vagin, A., Teplyakov, A., 1997. MOLREP: an automated program for molecular replacement. J. Appl. Crystallogr. 30, 1022–1025.

Villar-Palasi, C., Larner, J., 1970. Glycogen metabolism and glycolytic enzymes. Annu. Rev. Biochem. 39, 639–672.

Vonrhein, C., Blanc, E., Roversi, P., Bricogne, G., 2007. Automated structure solution with autoSHARP. Methods Mol. Biol. 364, 215–230.

Whitehouse, D.B., Tomkins, J., Lovegrove, J.U., Hopkinson, D.A., McMillan, W.O., 1998. A phylogenetic approach to the identification of phosphoglucomutase genes. Mol. Biol. Evol. 15, 456–462.

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., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242.

Chapter 3

STRUCTURAL STUDY OF MEMBRANE ENZYME

IPCT/DIPPS

NOTE: Project performed by Przemyslaw Nogly, supervised by Margarida Archer – head of Membrane Protein Crystallography Laboratory (ITQB-UNL, Oeiras) and help provided by Pikyee Ma. This work was performed in collaboration with the Cell Physiology & NMR Laboratory, headed by Helena Santos (ITQB-UNL, Oeiras) with her team Ana M. Esteves and Nuno Borges; Membrane Transporters Group, team of Valentin Gordeliy (IBS, Grenoble); and Isabel Moraes, Membrane Protein Laboratory, (Diamond Light Source, Didcot).

P.N. cloned, expressed, purified, optimized sample, crystallized the protein in detergent, optimized diffraction, measured X-ray diffraction data, and phased data at low resolution. I.M. supported crystallization and diffraction tests. P.N performed initial crystallization in lipidic sponge phase; V.G. team crystallized the protein in mesophases, collected data and solved the structure. P.M and P.N. performed mutagenesis. A.E., N.B. and H.S. performed the activity assays. M.A., P.N. and V.G. analyzed the structure and wrote the manuscript.

3.1 INTRODUCTION

Hyperthermophiles are able to withstand extreme temperatures and there is growing attention into the mechanisms they developed to survive thermal stress (Empadinhas and Da Costa, 2006; Trivedi et al., 2006). This knowledge may have potential applications as natural strategies to preserve activity of biomolecules will be very useful in both, medical industry and academic research, often aiming at unstable but of high interest human proteins (Mascellani et al., 2007). Moreover, fundamental mechanisms developed by organisms to adapt to very diverse environments is fascinating and still remains an open topic (Siddiqui and Thomas, 2008; Stetter, 1999). In recent phylogenetic analysis, Gonçalves et al. studied the evolution of the biosynthesis of di-myo-inositol-1,1´-phosphate (DIP), a marker of adaptation to hot environments (Gonçalves et al., 2012). DIP was found exclusively in a number of (hyper)thermophiles (Santos et al., 2011). The genes involved in the biosynthesis of DIP were identified (Borges et al., 2006; Rodrigues et al., 2007) and the complete biosynthetic pathway was reconstituted *in vitro* (Rodionov et al., 2007). Our attention was particularly attracted by the novel enzyme, CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) / CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (also named di-*myo*-inositol phosphate phosphate synthase and thus abbreviated DIPPS), that has two consecutive activities playing a key role in the biosynthesis of DIP. In most organisms, both genes coding for IPCT

and DIPPS activities are fused in a single gene, although separate genes are also found in a few organisms (Gonçalves et al., 2012). The bifunctional enzyme IPCT/DIPPS from *Archaeoglobus fulgidus* contains two domains with high homology to nucleotidyltransferases (N-terminal IPCT) and CDP-alcohol phosphatidyltransferases (C-terminal DIPPS). The recently published crystal structure of the cytoplasmic IPCT domain (Brito et al., 2011) shows the structural features typical of nucleotidyltransferase family (PF00483) and provides insights into the enzyme specificity for CTP substrate over other nucleoside triphosphates. However, structural characterization of the second domain, DIPPS, was hindered due to presence of predicted transmembrane helices and inherent difficulties in its expression and purification. In the present work we confirmed localization of the protein in the membrane fraction. The full length bifunctional protein was successfully overexpressed in *Escherichia coli* and we established the first protocol for IPCT/DIPPS purification. Pure and homogeneous protein of the full length IPCT/DIPPS was obtained allowing for crystallization experiments. Crystals were grown in crystallization with detergents diffracting to 4.5 Å. Moreover, the use of lipidic phases for crystallization resulted in crystals diffracting beyond 3 Å resolution and enabled solving the enzyme structure. To our knowledge this is the first report of crystallization and structure of a member of CDP-alcohol phosphatidyltransferase family (PF01066), which is a large family of

integral membrane proteins, including a number of enzymes involved in lipids biosynthesis and present in all domains of life.

3.2 PRODUCTION AND CRYSTALLIZATION OF IPCT/DIPPS

3.2.1 MATERIAL AND METHODS

3.2.1.1 CLONING, EXPRESSION AND PURIFICATION

The gene AF_0263 was amplified from *Archaeoglobus fulgidus* DSM 4304 chromosomal DNA using forward primer TGCCCCCGGGATGATAAATGTTGACGGAGAATAC and reverse primer TATGAGCTCTTTAGAAACCAAAACAGCAAGTAAAG. The gene was cloned into pET52b(+) vector using XmaI and SacI restriction sites which incorporate an N-terminal StrepII-tag and C-terminal 10xHis-tag into the expressed protein. The construct for expression starts with the protein sequence MINVDG, where M corresponds to 55th residue of the translated AF_0263 gene. The full sequence of the expressed gene is given in the Appendix 1.

The protein was expressed in *E. coli* strain C43 DE3 grown in terrific broth (TB) medium. On the day of expression 1% of an overnight culture was used to inoculate fresh medium with 100 µg ml^{-1} ampicillin. The cells were grown at 37°C with mixing at 200 rpm. When the OD_{600nm} reached 1.0 (in about 3 – 3.5 h) the temperature was lowered to 30°C and expression of the recombinant protein was induced by addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG). Additional 50 µg ml^{-1} of ampicillin was added 3 h post-induction and cells were harvested 6 h post-induction

with OD_{600nm} often reaching over 4.0. The biomass was harvested by centrifugation at 4000*g* for 20 min. The cell pellet was frozen in liquid nitrogen and stored at -80°C. If the growth was performed in a fermentor then pH was controlled at 7.4, in addition to the phosphate buffer already present in the TB medium.

All the steps of cell disruption and purification were performed at 4°C. In the usual purification the cells were disrupted in continuous flow cell disrupter in a buffer containing 50 m*M* sodium citrate, 50 m*M* sodium phosphate, 200 m*M* NaCl, 5% glycerol, 100 m*M* sucrose, protease inhibitor cocktail tablet (SIGMAFAST from Sigma-Aldrich, use following the supplier recommendation), 1 m*M* ethylenediaminetetraacetic acid (EDTA), 0.5 m*M* phenylmethanesulfonylfluoride (PMSF), 2 m*M* β-mercaptoethanol (final pH 7.5 at 4°C). *E. coli* membranes were isolated from cleared cell lysate (30 min centrifugation at 10000*g*) obtained from 30 l fermentation by ultracentrifugation at 200000*g* for 2 h. The protein was solubilized from resuspended *E. coli* membranes in the final volume of 200 ml by gentle stirring at 4°C for 2 h in the presence of 0.7% (w/v) DDM (Anagrade; Affymetrix) or 2% (w/v) Triton X-100 (protein grade; Merck). Besides detergent the solution used for solubilization contained 20% glycerol, 1 *M* NaCl, 50 m*M* phosphate buffer, 50 m*M* imidazole, 100 m*M* sucrose, 2 m*M* β-mercaptoethanol and 1 m*M* $MgCl_2$ (final pH 7.4 at 4°C). The insoluble fraction after solubilization was separated by 1 h centrifugation at 100000*g*. The supernatant was loaded on three 5 ml HisTrap HP columns (GE

Healthcare) connected together. After loading the sample the columns were extensively washed with 20 – 40 column volumes of 50 m*M* phosphate buffer, 50 m*M* imidazole, 10% glycerol, 1 *M* NaCl (final pH 7.4 at 4°C). In the following washing step with 110 m*M* imidazole (otherwise buffer composition as in the wash buffer) was carried out with about 10 column volumes or until absorbance at 280 nm reached baseline. The protein of interest was eluted in a step elution with 500 m*M* imidazole, 50 m*M* phosphate buffer, 10% glycerol, 300 m*M* NaCl, 1 m*M* $MgCl_2$ (final pH 7.4 at 4°C). Immediately after elution the sample was supplemented with 1 m*M* β-mercaptoethanol. The washing and elution buffers contained additionally one of the following detergents: 0.1 - 0.2% Triton X-100 (Merck, protein grade), 0.05% DDM (Affymetrix, Anagrade). The sample was concentrated to 20 mg ml^{-1} in centrifugal filter unit with membrane of 100 kDa molecular weight cut-off. The protein behaves well during the concentrating step as long as the buffer contains at least 300 m*M* NaCl. In some cases the buffer was exchanged in the size exclusion chromatography to 10 m*M* CHES pH 9.5, 500 m*M* NaCl, 1 m*M* $MgCl_2$, 1m*M* β-mercaptoethanol and the respective detergent used in purification.

When protein sample was needed in the presence of a different detergent than the one used for membrane solubilization, the detergent was exchanged while the protein was bound to HisTrap column and washed in buffer containing the detergent of choice.

3.2.1.2 HOMOGENEITY TESTS

Homogeneity tests were performed by diluting 10 times the initial protein sample at 27.5 mg ml^{-1} (with 0.5% Cymal-5 i.e. 2 times the critical micelle concentration, CMC) with variety of detergents at concentrations about two times higher than their CMC. The buffer besides the detergents tested contained 20 m*M* CHES pH 9.5, 500 m*M* NaCl, 10% glycerol, 1 m*M* $MgCl_2$, 1 m*M* tris(2-carboxyethyl)phosphine (TCEP). The samples were incubated at 20°C for 6 days followed by size exclusion chromatography (Superdex 200 5/150 GL, GE Healthcare) to assess the effect of the detergents on the sample homogeneity. It is worth to note that the incubated samples in addition to the tested detergent contained also Cymal-5 from the initial protein sample at 0.05%, which corresponds to concentration five times lower than its CMC. Thus, the protein was incubated in a mixture of detergents. Some of the samples showed precipitation after 6 days incubation at 20°C, so these samples were centrifuged and the supernatant was also analyzed by SEC.

3.2.1.3 CRYSTALLIZATION "IN DETERGENT"

Initial screening of the crystallization conditions was performed with Cartesian Mini-Bee nano-robot at 20°C using 13 crystallization screenings (over 1248 conditions). The vapor-diffusion crystallization with sitting drops was prepared by mixing 200 nl of protein sample (at concentration 20 mg ml^{-1} in buffer containing 500 m*M* imidazole, 50 m*M* phosphate buffer, 10% glycerol, 300 m*M* NaCl, 1 m*M* $MgCl_2$,

1m*M* β-mercaptoethanol and tested detergent) and 200 nl of reservoir solution. For each crystallization condition three crystallization drops were set up, each one containing protein sample in the presence of different detergent. The detergents tested initially included DDM (0.025%), Triton X-100 (0.25%), DM (0.2%) and OG (0.8% in mixture with mixture with 0.16% Triton X-100).

3.2.1.4 DATA COLLECTION AND PROCESSING

The weakly diffracting IPCT/DIPPS crystals obtained in the presence of Triton X-100 were measured at Swiss Light Source (SLS) beamline PXIII. Data were processed with XDS (Kabsch, 2010) and scaled with Scala (Evans, 2011). Molecular replacement was performed with Phaser (McCoy et al., 2007). The details are presented in Table 3.2.

3.2.2 RESULTS AND DISCUSSION

The IPCT/DIPPS was cloned into vector pET52b(+) and expressed in *E. coli* C43 DE3 strain grown in terrific broth medium. The *E. coli* strain C43 DE3 gave better expression as compared with other tested strains such as C41 DE3, Rosetta and Bl21 DE3. Increased expression level was obtained when *E. coli* C43 DE3 cells were transformed with the additional plasmid encoding for rare *E. coli* codons. However higher expression yield was accompanied by extensive protein degradation and thus the latter was not used.

The full length IPCT/DIPPS was purified from *E. coli* membranes. The protein was solubilized in Triton X-100 or DDM and purified using immobilized metal ion affinity chromatography. Good purity of the protein was obtained by minimizing the ionic interactions with addition of high salt concentration (1 *M* NaCl) in both loading and wash buffer of immobilized metal ion affinity chromatography. Importantly, the protein construct with the deca-histidine tag resulted in a strong binding of the protein to the HisTrap column allowing removal of contaminating proteins with 110 m*M* imidazole (Figure 3.1). The protein was eluted in one step, increasing the imidazole concentration from 110 m*M* to 500 m*M*. The usual protein yield of purification was 2 – 4 mg per one liter of growth medium.

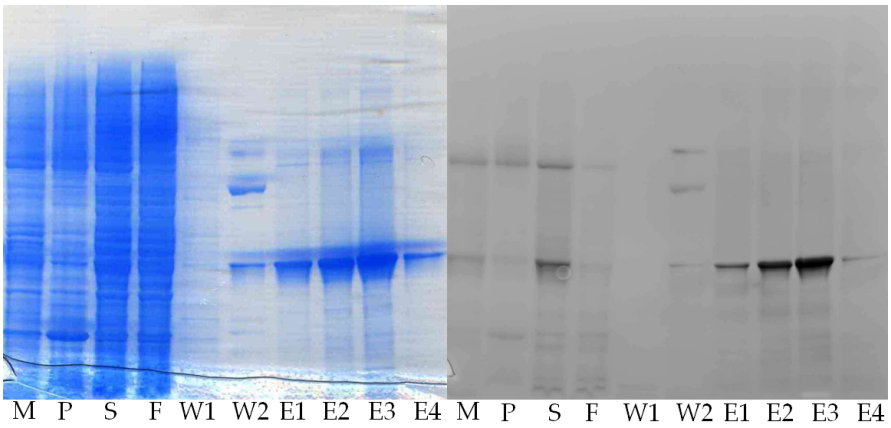


Figure 3.1 Coomassie blue stained SDS-PAGE of fractions from purification of IPCT/DIPPS (left) and His-tag specific InVision, Invitrogen (right). Lanes contain fractions: *M* – membrane, *P* – pellet after membrane solubilization, *S* – sample applied on HisTrap, *F* – unbound fraction, *W1* - wash with 50 mM imidazole, *W2* – wash with 110 m*M* imidazole, *E1-4* – eluted fractions with 500 m*M* imidazole.

Optionally, a size exclusion chromatography was used as the final polishing step of purification. Depending on the batch of purification 0 - 50% of the protein elutes in the void volume. However, the pool of non-aggregated fractions is stable and homogenous over time.

Detergents forming small micelles are assumed to expose more of the hydrophilic regions of membrane proteins which are predominantly involved in formation of crystal contacts (Bill et al., 2011; Kunji et al., 2008). Thus smaller detergent micelles may promote formation of well ordered crystals (Ostermeier and Michel, 1997; Stroud, 2011). On the other hand the smaller detergent micelles are known to stabilize membrane proteins less effectively, leading to protein aggregation and precipitation. Monodispersity of the protein is often mentioned as one of the important parameters promoting crystallization (Berry et al., 2006; Geerlof et al., 2006; Proteau et al., 2010).

In order to identify the detergents that effectively stabilize the protein over time (required for its crystallization) but also form small micelles, the impact of thirty detergents on IPCT/DIPPS monodispersity was tested by SEC (Table 3.1). Six out of thirty protein samples precipitated after incubation at 20°C for six days. However, the samples of protein solubilized with n-octyl-β-D-thioglucopyranoside (OTG) and n-octanoylsucrose, after spinning down the precipitation, appear to be only minimally aggregated. Thus it is not clear whether the observed white precipitate is due to protein precipitation or a result of progressive detergent degradation. In contrast, the samples that precipitated with octanoyl-N-

methylglucamide (MEGA-8), cyclohexylpropanoyl-N-hydroxyethylglucamide (C-HEGA-9), octanoyl-N-hydroxyethylglucamide (HEGA-8), did not show any absorbance at 280 nm after SEC, an indication that all the protein had precipitated.

The protein appears to be monodisperse in the detergents commonly used for membrane protein purification, as for example n-dodecyl β-D-maltoside (DDM) and Triton X-100. Among the detergents forming small micelles and stabilizing the IPCT/DIPPS protein, are lauryldimethylamine-N-oxide (LDAO) and n-decyl β-D-maltoside (DM). In addition, two of the detergent series with cyclic hydrophobic chain, cyclohexyl-1-pentyl-β-D-maltoside (Cymal-5) and 4-cyclohexyl-1-butyl-β-D-maltoside (Cymal-4) stabilize the protein well. Both detergents are characterized by small micelles and may also be suitable for crystallization. Interestingly, the detergents of the glucoside and maltoside detergent series with short hydrocarbon chains formed by 9 or 8 carbon atoms appear to be less effective than their counterparts in the thioglucoside and thiomaltoside series, which are also characterized with lower CMC values.

Table 3.1 Monodispersity tests of IPCT/DIPPS in various detergents. Samples were incubated for six days at 20°C and analyzed by size exclusion chromatography (SEC). Classification of aggregation level: none, <10% (*), 10 – 50% (**), >50% (***). The precipitated samples were centrifuged and supernatant was also analyzed by SEC; if no absorbance at 280 nm was observed, it is assumed that the entire protein sample precipitated (-). The SEC profiles and the abbreviations of detergent names are included in Appendix 3.

Detergent	After 6 days: precipitation (P), clear (C)	Aggregation
DDM	C	*
Triton X-100	C	none
DM	C	none
LDAO	C	none
Cymal-5	C	none
NG	C	**
OTG	P	*
OG	C	***
Cymal-3	C	*
MEGA-8	P	-
CHAPS	C	*
CHAPSO	C	none
C-HEGA-9	P	-
HEGA-8	P	-
C-HEGA-8	C	*
DTM	C	*
Cymal-4	C	none
OTM	C	*
Fos-Choline-10	C	none
Fos-Choline-9	C	***
Fos-Choline-8	C	***
C12E8	C	*
ZWITTERGENT 3-12	C	*
DDAO	C	**
n-Octanoylsucrose	P	*
ZWITTERGENT 3-10	C	***
HEGA-10	C	*
C8E5	C	***
NTG	C	*
HTG	P	***

The initial crystallization trials yielded a hit with the nuclear receptor - ligand binding domain crystallization screen (Molecular Dimensions) in the A5 condition containing 9% PEG 8000, 100 m*M* Tris pH 8.0 (Figure 3.2a). The protein sample used in this screening was applied directly after elution from the HisTrap column in the presence of 0.2% Triton X-100 and addition of the 1% n-Octyl-β-D-Glucopyranoside. The 15 µm crystals diffracted very weakly at the European Synchrotron Radiation Facility (ESRF) synchrotron source at 30 Å (Figure 3.2a). Upon extensive optimization of the crystallization conditions, hexagonal plate crystals and similar to hexagonal bipiramides, which however have cut off the two apices, were obtained with IPCT/DIPPS protein sample solubilized in Triton X-100 (Figure 3.2b-d). These crystals were obtained by vapor-diffusion method in hanging drop with reservoir solution composed of 18% PEG 5000MME, 1 *M* NaCl, 100 m*M* BisTris pH 6.5 and mixed with a 1:1 ratio (volume) with protein sample. Larger crystals, up to 300 µm were grown in drops where additionally one volume of water was added to initial drop formed with one volume of protein and one volume of reservoir. Interestingly, other IPCT/DIPPS batches showing some protein aggregation and not subjected to SEC purification still yielded crystals of similar diffraction quality. However, the diffraction of the crystals in all cases varied significantly even within the same crystallization drop. The crucial factors for crystallization appear to be high NaCl concentration. In the presence NaCl at concentrations lower than 700 m*M* in the

crystallization condition the crystals diffraction become worse. While higher NaCl concentrations decrease nucleation and improve diffraction. On the other hand concentration of PEG 5000MME over 20% appears to deteriorate the resolution limit to which crystals diffract. Addition of DIP at 25-50 m*M* into the crystallization drop was also used to yield better diffracting crystals.

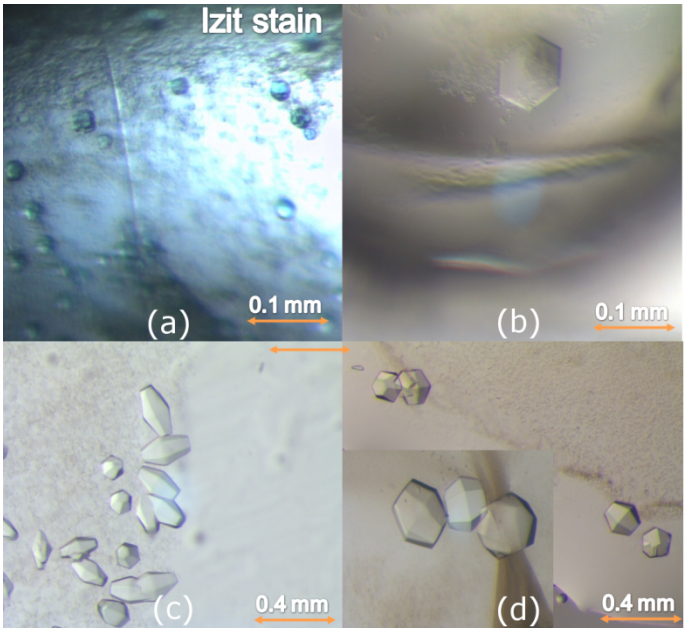


Figure 3.2 Crystals of IPCT/DIPPS obtained in the presence of Triton X-100 (examples). (a) Initial crystals obtained in the presence of PEG 8000. (b-d) Crystals obtained upon optimization in the presence of PEG 5000MME.

Crystals obtained in PEG 5000MME (Figure 3.2c-d) diffracted up to 4.5 Å at synchrotron sources but with anisotropic diffraction and often not sufficient quality to be processed successfully. One dataset was processed up to 6.6 Å (Table 3.2).

Section 3.3.1 of this chapter describes crystallization in lipidic mesophases and structure solution of IPCT/DIPPS. This structure was also used as a search model for molecular replacement phasing of data collected at 6.6 Å from crystals obtained in the presence of detergent described in this section. The protein in the presence of detergent crystallized in the hexagonal space group $P6_522$ with two molecules in the asymmetric unit, while the crystals grown in lipidic phases belonged to the orthorhombic space group $P2_12_12$ with one molecule in the asymmetric unit (Table 3.2). In the presence of two IPCT/DIPPS molecules in the asymmetric unit the solvent content of the crystal was 73.4%.

The molecular replacement solution has no clashes in the crystals packing. Moreover, a continuous electron density for the secondary structure elements is observed, including the transmembrane helices of DIPPS domain (Figure 3.3). However, the low resolution of this data set hinders refinement of the structure.

The obtained solution indicates that the overall interdomain geometry of IPCT/DIPPS is preserved in the two different methods of membrane proteins crystallization, in lipidic mesophases and in detergent. The dimer interface is formed through the DIPPS domain.

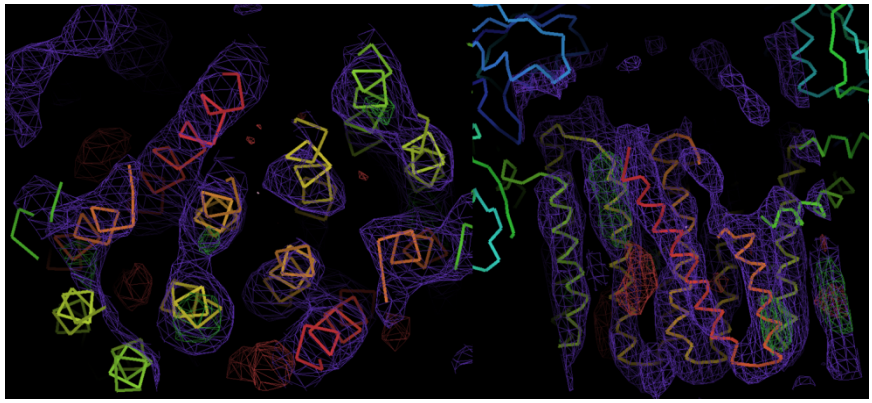


Figure 3.3 Molecular replacement solution with data collected from crystal obtained in crystallization with Triton X-100. The protein polypeptide chain is represented by C^{α} atoms. Shown fragment of the DIPPS transmembrane domain illustrates the continuous electron density for the secondary structure elements in the view parallel to the expected membrane surface (left) and across the expected membrane (right). The data used in molecular replacement was at 6.6 Å resolution and the electron density cut-off shown in the figure is at 1 RMSD.

3.3 STRUCTURE DETERMINATION OF THE BIFUNCTIONAL ENZYME IPCT/DIPPS

3.3.1 MATERIALS AND METHODS

3.3.1.1 CRYSTALLIZATION IN LIPIDIC PHASES

The sample of IPCT/DIPPS was subjected to lipidic sponge phases crystallization screening (Molecular Dimensions) with 48 different conditions following the manufacture´s protocol. The crystallization drops were equilibrated against the solution which yielded crystals in detergent (described in section 3.2.2) and in one trial this solution was also added to the crystallization drop accounting for about 30% of its volume. Crystals were obtained in about 16 out of the 48 conditions of the screening indicating that lipidic phases may be a suitable medium for crystallization of IPCT/DIPPS (Figure 3.4).

Protocols for further crystallization screening and up-scaling of the crystallization in lipidic mesophases were carried out by the group of Valentin Gordeliy (IBS, Grenoble). These efforts resulted in crystals of 100 µm that diffracted beyond 3 Å.

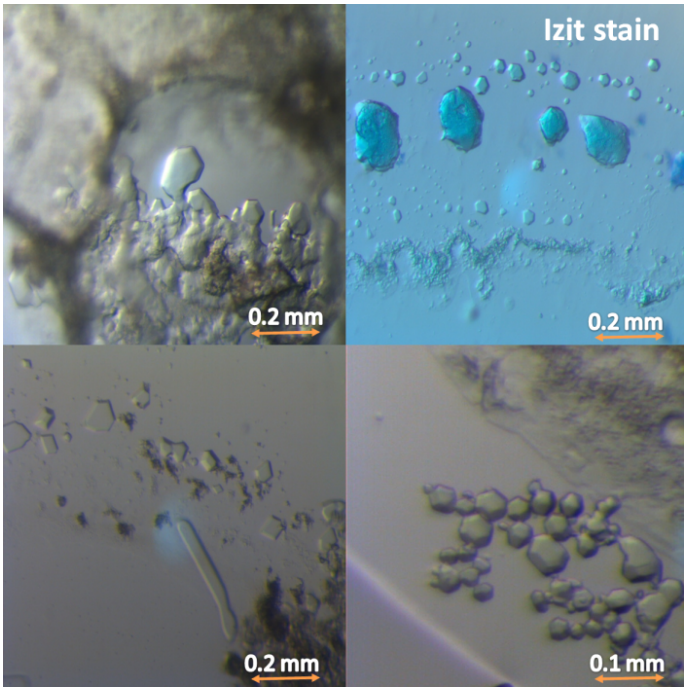


Figure 3.4 Examples of IPCT/DIPPS crystals obtained in some conditions of the lipidic-sponge phase crystallization screen (Molecular Dimensions).

3.3.1.2 DATA COLLECTION AND PROCESSING

Diffraction data were collected from crystals obtained in lipidic mesophases at ID23-1 and ID23-2 beamlines of the European Synchrotron Radiation facility (ESRF, Grenoble, France). The data were integrated to 2.85 Å (R_{merge} at the border is 59%, $I/\sigma(I)$ is 1.9). Data collection and processing, phasing and crystallographic refinement were performed by the group of Valentin Gordeliy (Table 3.2).

Table 3.2 Data collection and processing statistics of IPCT/DIPPS.

Crystallization method	"In detergent"	Lipidic mesophase
X-ray source	SLS beamline, PXIII	
Wavelength (Å)	1.000	
Temperature (K)	100	
Detector	PILATUS 2M-F	
Rotation per image (°)	0.2	
Space group	$P6_522$	$P2_12_12$
Unit-cell parameters (Å)	a = b = 107.53, c = 609.23	a=41.33, b=107.65, c= 123.22
Resolution range (Å)	54.72 - 6.60 (6.96 – 6.60)	40.67 - 2.852
Total No. of reflections	72653 (10724)	
No. of unique reflections	4554 (630)	13318 (2415)
Completeness (%)	99.8 (100.0)	97.93 (91.0)
Average multiplicity	16.0 (17.0)	
Average I/σ(I)	8.2 (1.6)	
R_{pim} (%)	5.0 (46.0)	
Overall B factor (Å^2)	-	60.98
Statistics on refinement		
Resolution (Å)	-	2.85 (3.07-2.85)
No. reflections	-	13318
R_{work} / R_{free} (%)	-	24.7/30.6
R.m.s. deviations		
Bond lengths (Å)	-	0.003
Bond angles (°)	-	0.631

Values for the higher resolution shell are given in parentheses.

3.3.1.3 PHASING AND CRYSTALLOGRAPHIC REFINEMENT

Initial phases were obtained by molecular replacement using the crystallographic structure of the IPCT soluble domain (PDB ID 2XME, Brito et al. 2011) as a starting model. Search for α-helical fragments resulted in establishing the position of five transmembrane α-helices. Refinement of the structure fragments improved the electron density maps, allowing the identification of the sixth transmembrane helix and two α-helices parallel to the membrane plane. The crystallographic model was further refined until final R-factors of 24.7% (R_{free} of 30.6%). The structure contains all the residues of the transmembrane domain except for the disordered loops (291 - 298 and 430 - 440) and C-terminal lysine (numbering following UniProt entry O29976).

3.3.1.4 MUTAGENESIS AND ACTIVITY TESTS

After obtaining the structure of IPCT/DIPPS, mutagenesis were attempted in order to elucidate the role of amino acid residues within the putative active site and potentially involved in substrate binding. Primers for mutagenesis were designed using QuikChange primer design web tool. The point mutations were introduced using the QuikChange I site-directed mutagenesis kit (Agilent Technologies). All point mutations were confirmed by sequencing using the StabVida sequencing service (Caparica, Portugal). The mutants were cultured following the protocol described for the wild type (section 3.2.1.1).

The activity tests were performed by team of Prof. Helena Santos from Cell Physiology & NMR Laboratory (ITQB-UNL, Oeiras) including Ana M. Esteves and Nuno Borges. Cells were resuspended in 50 m*M* Bis-Tris propane (pH 8.0) containing 5% glycerol and 100 m*M* NaCl, and disrupted in a French press. Cell debris was removed by centrifugation (10000*g*, 15 min, 4°C) and the supernatant was used to measure IPCT/DIPPS activities. Reaction mixtures (0.3 ml) containing 25 m*M* Bis-Tris propane (pH 8.0), 10 m*M* $MgCl_2$, 2.7 m*M* CTP, 7 m*M* L-*myo*-inositol 1-phosphate, and cell extract (0.9 mg of total protein) were incubated at 85°C and samples were collected after 30 minutes. The reactions were stopped by placing the sample on ice. The metal ions were chelated with ethylenediaminetetraacetic acid (EDTA) and the amounts of the product and reaction intermediate were assessed using ^{31}P NMR using diglycerol-phosphate as an internal concentration standard.

3.3.2 RESULTS AND DISCUSSION

3.3.2.1 X-RAY STRUCTURE OF IPCT/DIPPS

The final model comprises one IPCT/DIPPS molecule with 396 amino acid residues (from one polypeptide chain), four lipophilic fragments of lipids and one Mg^{2+}. The IPCT/DIPPS molecule is composed of two domains (Figure 3.5).

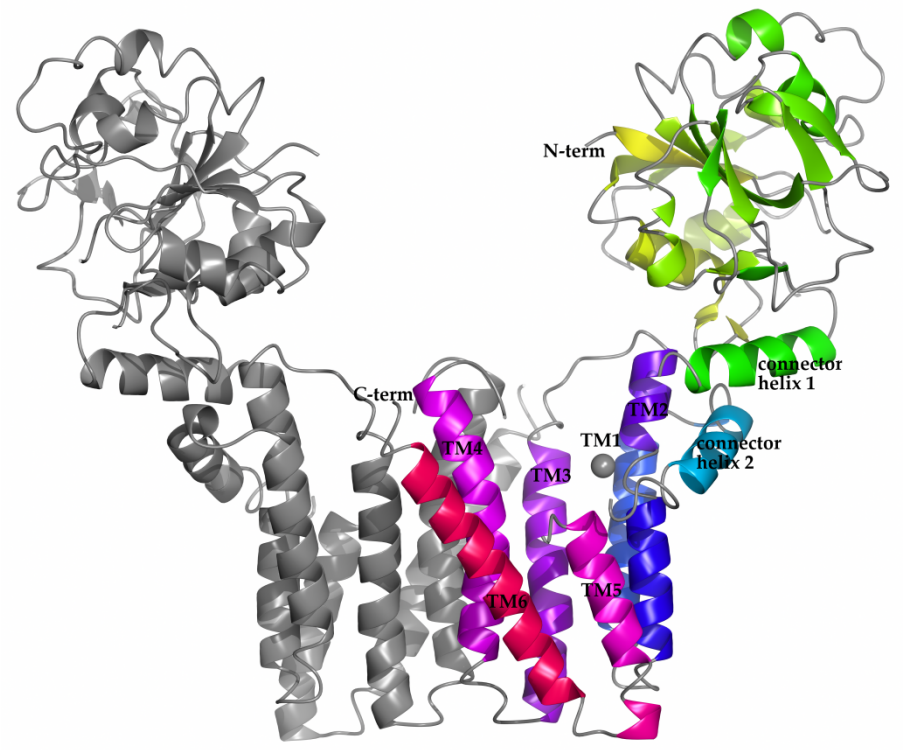


Figure 3.5 IPCT/DIPPS dimer. Secondary structure shown in ribbon representation. One monomer colored; yellow to green – IPCT, cyan to red – DIPPS, and the related by crystallographic symmetry polypeptide chain A´ in grey. The labels indicate two connector helices and six transmembrane helices (TM1-TM6); the grey sphere indicates putative Mg^{2+} ion. The secondary structure elements were assigned with DSSP (Kabsch and Sander, 1983). This figure and others in this chapter were drawn with CCP4MG (McNicholas et al., 2011) unless stated otherwise.

The cytoplasmic IPCT domain (Figure 3.6a), as described in detail by Brito et al. (2011), has a fold typical of cytidylyltransferases somewhat similar to the dinucleotide-binding Rossmann fold (Rao and Rossmann, 1973). The central part consists of a seven-stranded mixed β-sheet (β1, β4, β5, β6, β8, β12 and β11). It is flanked on one side by a two-stranded antiparallel β-sheet β7 and β13 (which together with β2 and β3 form an interface to the connector helix 1) and helices α1, α2, α4. On the other side, where the active site is located, there is a two-stranded antiparallel β-sheet, β9 and β10, and helices α3, α5, α6. The IPCT domain of the bifunctional enzyme structure is very similar to the previously published X-ray structure of the IPCT domain alone determined at 1.9 Å resolution (Brito et al., 2011), with a RMSD for all C^{α} atoms of 0.6 Å. The C-terminal of the IPCT alone was truncated at amino acid residue 286 and the crystallographic model of IPCT lacks the last 10-12 residues. The major difference is the inclusion of the connector helix 1 in the bifunctional construct, which covers the hydrophobic face of the β-sheet formed by β7 and β13. This connector helix 1 prevents the interactions mediated through this short β-sheet observed in the truncated form of IPCT, which led to the formation of an apparent dimer.

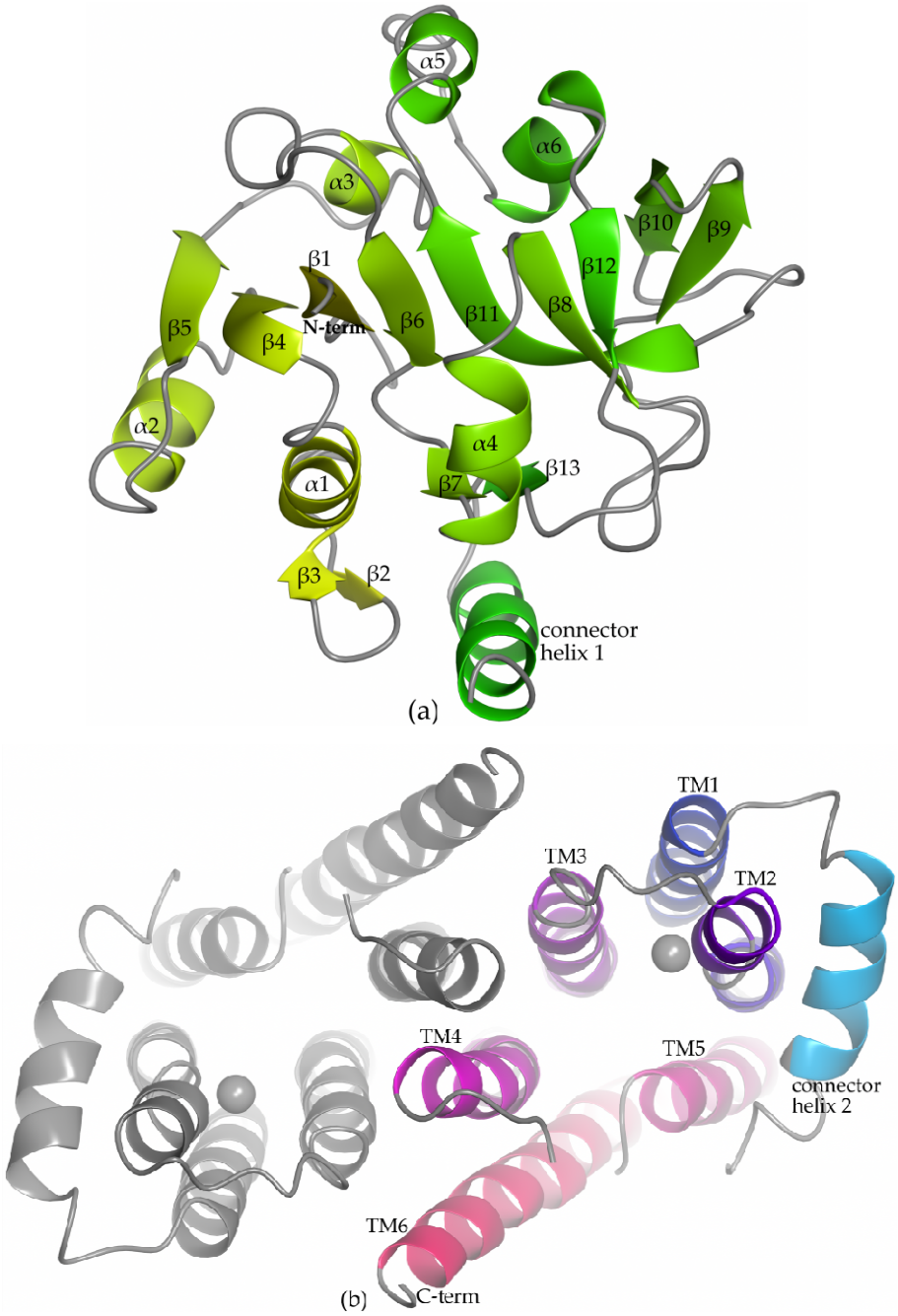


Figure 3.6 Secondary structure representation of (a) IPCT domain and (b) dimer formed by DIPPS domains (view perpendicular to the membrane plane).

The cytoplasmic IPCT domain is connected to the membrane DIPPS domain via two helices nearly parallel to the membrane (connector 1 and 2) and a flexible loop (Figure 3.5). DIPPS domain comprises six transmembrane helices (Figure 3.6b), four of which (TM2, TM3, TM4, TM5) resemble a four-helix bundle fold (Kamtekar and Hecht, 1995). The transmembrane helices and connector helix 2 form a hydrophilic cavity in the central part of the domain, which widely opens towards the cytoplasm. One of the loops (between TM2 and TM3) located at the entrance of this cavity comprises amino acids 369-373, and in spite of the high thermal motion parameters shown by these residues (average B-factor for C^{α} atoms is 84 $Å^2$), it is visible in the electron density maps. In contrast, the two additional loops that should also be placed at the entrance could not be modeled, likely due to their flexibility. One of the loops (residues 291 - 298) is situated between the connector helix 1 and connector helix 2, and the other loop (residues 430 - 440) connects TM4 and TM5.

3.3.2.2 QUATERNARY STRUCTURE

The IPCT/DIPPS molecule, present in the crystal asymmetric unit, forms extensive contacts with another molecule related by crystallographic symmetry (operator -x+2,-y+1,z) (Figure 3.5). This interface is entirely formed by interactions between the two DIPPS domains, involving mainly the hydrophobic residues sited at TM3, TM4 and TM6 (Figure 3.7). Further stabilization is achieved by 8 hydrogen bonds and 2 salt bridges established among residues

located near the cytoplasmic membrane surface and in the extramembrane region. The interface area extends over a large surface of 1936 $Å^2$ and the CSS value of 1 assigned by PISA implies that the interface has an essential role in complex formation (Krissinel and Henrick, 2007).

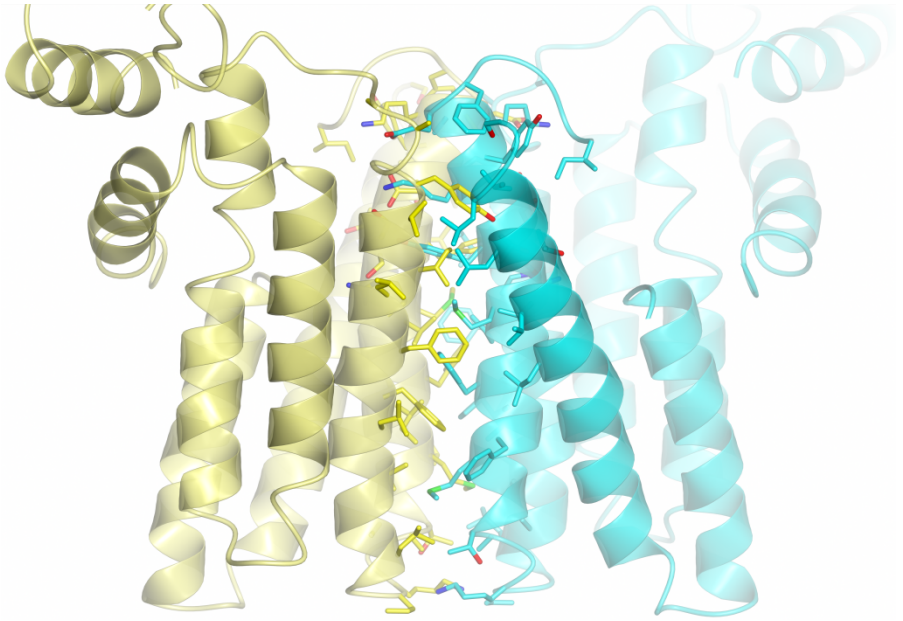


Figure 3.7 IPCT/DIPPS dimer interface. Residues involved in the interface formation shown in sticks representation. One IPCT/DIPSS molecule is displayed in cyan, while the symmetry-related molecule is colored yellow. C atoms have the same color as the molecule they belong to. The other atoms are colored: O – red, S – green, N – blue.

3.3.2.3 INTERFACE BETWEEN IPCT AND DIPPS DOMAINS

IPCT and DIPPS domain are found as separate enzymes in some of hyperthermophilic prokaryotes, like *Thermotoga maritima* and *Pyrolobus fumarii* (Gonçalves et al., 2012). The fusion of these two activities in the bifunctional enzyme from *Archaeoglobus fulgidus* reported here raises an interesting question about possible cooperativity between the two domains catalyzing consecutive steps in the biosynthesis of DIP.

The connector helix 1 is an integral part of the IPCT domain and is also present in a number of other nucleotidyl transferases. Comparison with 25 similar structures from Dali search (Holm and Rosenström, 2010) indicates that the position of this helix is highly conserved (Figure 3.12), covering the hydrophobic surface of strands β2, β3 and β7, β13 from IPCT.

The crystal structure shows that IPCT domain binds to the solvent exposed edge of the DIPPS domain (Figure 3.8). Interactions at this interdomain region involves amino acid residues from connector helix 1, short two-stranded β-sheet (β2, β3) and the helix α1 (IPCT) with connector helix 2 and TM2 (DIPPS domain). The interface extends over only 479 Å^2 and is stabilized by a few hydrophobic contacts and 4 hydrogen bonds. The CSS score of 0 in PISA server for interface between IPCT and DIPPS implies that it can be easily dissociated and seems to be a result of crystal packing only. Additionally, low conservation of the residues involved in this interface formation may indicate its non-essential character.

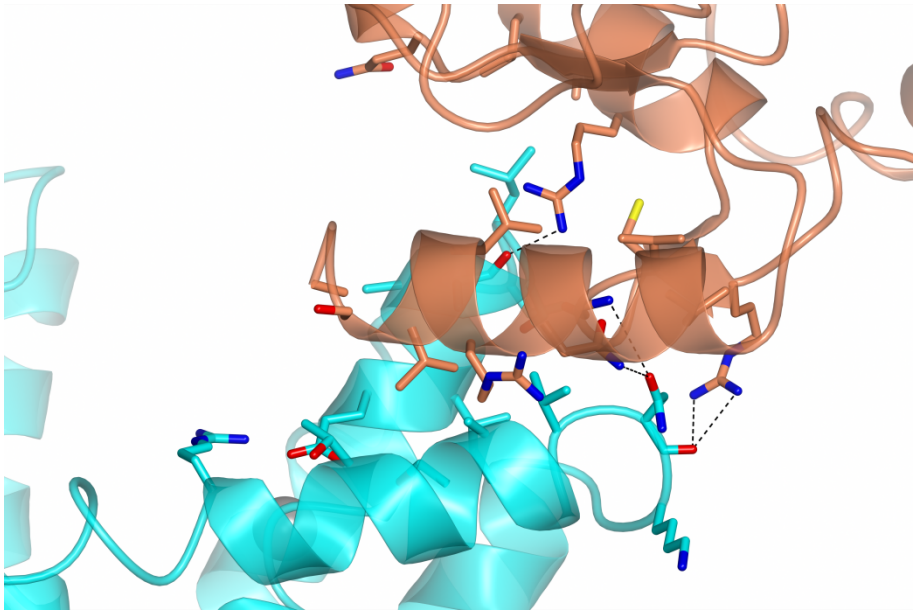


Figure 3.8 The interface between the IPCT and DIPPS domain. The residues involved in the interaction are represented as sticks. The C atoms and ribbons are colored cyan - DIPPS domain and coral - IPCT domain. H-bonds between domains indicated with dashed line. The other atoms are colored: O – red, S – yellow, N – blue.

3.3.2.4 POSITION OF DIPPS DOMAIN IN THE MEMBRANE

The DIPPS domain, comprising six transmembrane helices, forms an integral part of the cytoplasmic membrane. The position of these six α-helices in the membrane was predicted by PPM web server (Lomize et al., 2012) and is shown in Figure 3.9a. In this model all the extramembrane loops are localized on the cytoplasmic side of the membrane. In addition, the extended transmembrane helices TM1, TM2, TM4 and the C-terminus of TM6 are partially localized in the cytoplasm. Interestingly, TM2 has a kink introduced by Gly358, which coincide with the membrane surface.

The loops at the opposite side of the membrane appear to be embedded entirely inside the membrane, with predominantly hydrophobic residues and only a few charged ones in the proximity of the membrane surface, which could interact with the hydrophilic part of the phospholipid bilayer. This model seems to be well justified by the electrostatic surface of the domain shown in Figure 3.9b.

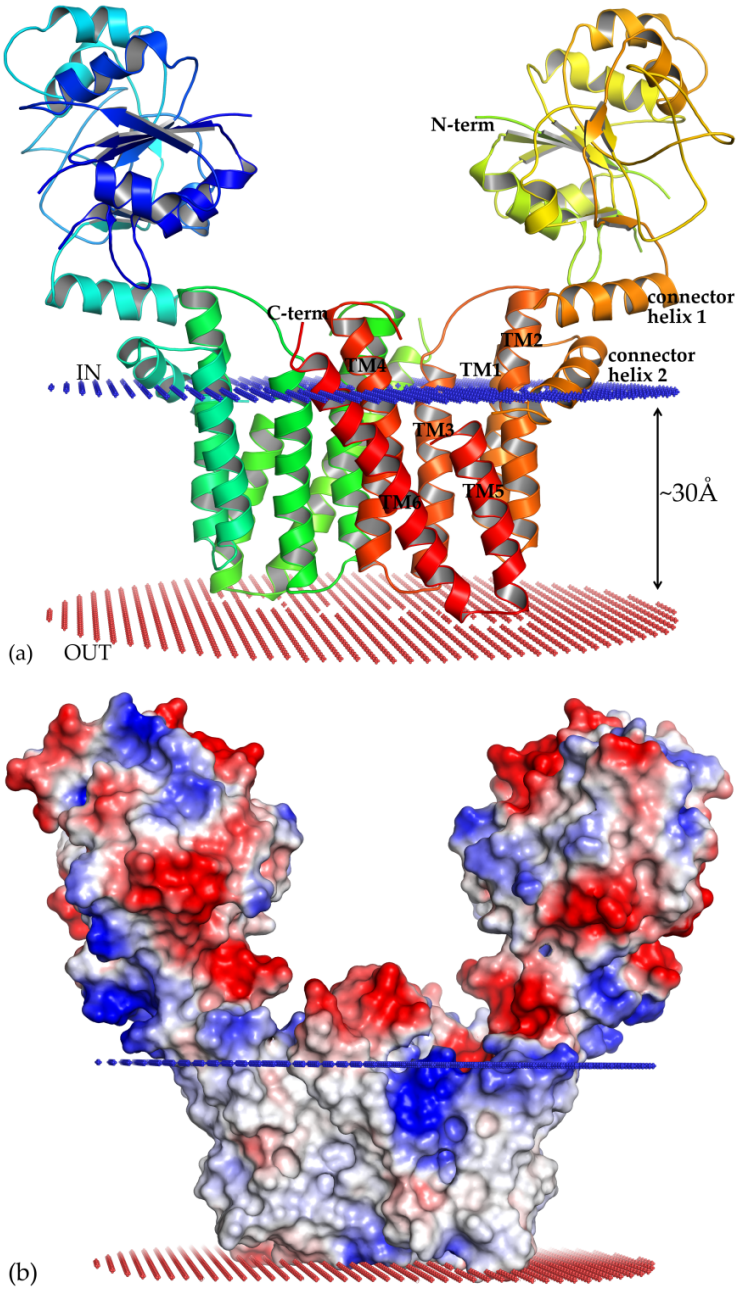


Figure 3.9 Predicted position of IPCT/DIPPS in the membrane (a) and corresponding electrostatic surface of the protein (b). IN and OUT labels indicate respectively cytoplasmic and extracellular side of the membrane. Molecular surface: negatively charged - red, positively charged - blue. The figures were draws using PyMOL and the position of the protein in the membrane was calculated using PPM server (Lomize et al., 2012).

Connector helix 1 has a partially amphipathic character, its hydrophobic side faces the hydrophobic surface of β-sheet (β7, β13) of IPCT, while the side-chains of positively charged residues are directed towards the surface of the membrane (Figure 3.10) with C^{α} distance to predicted surface of about 10 Å.

Connector helix 2 passes through the predicted surface of the membrane in a slightly sloped angle from parallel orientation. The beginning of this helix, from Ile299 to Ser308, is embedded inside the membrane and the positively charged residues of this segment are pointing towards the hydrophilic membrane surface.

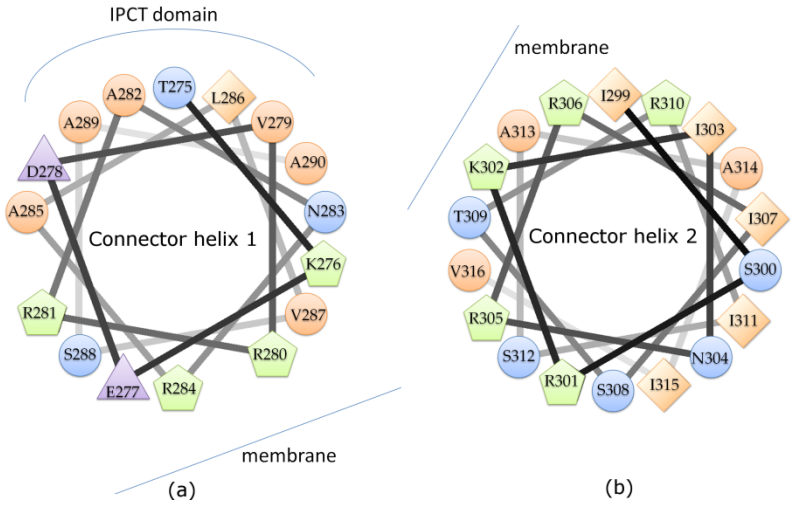


Figure 3.10 Helical wheel representation of connectors helix 1 (a) and helix 2 (b). Lines around the wheel indicate the approximate position of the helix with respect to the membrane. The membrane position of connector helix 2 (b) is representative for the fragment embedded in the membrane, Ile299 to Ser308.

A negatively charged patch is observed at the center of IPCT/DIPPS dimer (Figure 3.9b), formed by six glutamate residues at TM4 and inter-connecting loop TM4-TM5, (numbered 419, 424 and 427). The negative charge at the cytoplasmic end of TM4 protrudes large part of the helix out of the membrane into the cytoplasm and results in about 7 Å distance of TM4 (at Thr402) away from the calculated extracellular surface of the membrane.

Also the other side of the hydrophilic cavity scaffold formed by TM1 and TM2 (Figure 3.9a) protrudes from the membrane into the cytoplasm embraced by the amphipathic connector helix 2. This helix is predicted nearly parallel to the membrane surface and together with loop connecting to TM1 wraps around the TM2. The positions of TM1 and TM2 are 11.9 Å and 8.7 Å (at C^{α} of Leu337 and at C^{α} of Pro344) away from the calculated extracellular membrane surface.

3.3.2.5 CRYSTAL PACKING IN LIPIDS AND IN DETERGENT

The IPCT/DIPPS three-dimensional structure obtained from crystals grown in the lipidic phase is formed by layers of DIPPS domain, interconnected with layers of IPCT domains, which are responsible for the majority of the crystal contacts (Figure 3.11). The intermolecular contacts between the DIPPS dimers are limited to weak interactions. One involves residues at the end of connector helix 2 (Ala314, Ile315, Lys318) with the loop between TM5 and TM6 (Ala 458, Ile459) of the neighbor molecule, which form two hydrogen bonds and two hydrophobic contacts. Moreover, these hydrophobic interactions appear to be extended on both sides of the loop via two hydrocarbon chains (probably fragments of lipids) defined in the electron density. The other, involves hydrophobic interactions of Val481, Leu484, Ala485 and Val488, sited at the C-terminal ends of TM6 from two neighboring DIPPS dimers. .

Interactions of the IPCT domains are much more extensive in the crystal lattice. Each IPCT domain interacts with DIPPS domain of the same molecule (as described in section 3.3.2.3), three IPCT domains of the same layer and two IPCT domains of the neighboring layer. The strongest of these contacts is formed between IPCT domains of different layers. It comprises an interface area of 323.4 $Å^2$ and the amino acid residues involved in H-bonding are located at α5 and loops between α3 and β6, β9 and β10, α6 and β12.

In comparison with the crystal packing of crystals obtained in the presence of Triton X-100 (Figure 3.11; described in sections 3.2.1.3 and

3.2.2) the crystal lattice of crystals from lipidic phase appears much more ordered. In both cases, the lattices form layers of membrane domains, with the layers linked only via cytoplasmic domains. The distance between adjacent layers of membrane domains is around 80 Å in crystals from lipidic phase, and 46 Å (shortest distance) in crystals grown in the presence of Triton X-100, also showing a more irregular arrangement.

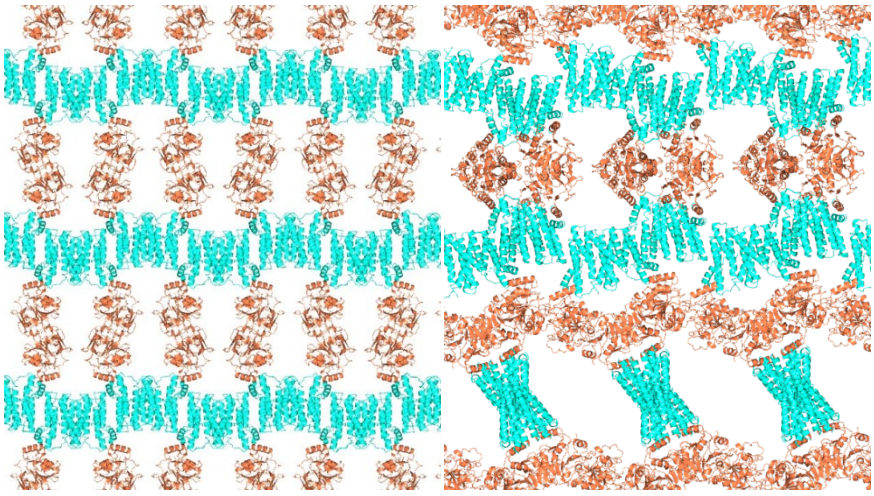


Figure 3.11 Crystal packing of crystals obtained in lipidic mesophases (left) and in the presence of Triton X-100 (right). The two domains are shown in different colors, IPCT – coral, DIPPS – cyan and the secondary structure represented as ribbons.

3.3.2.6 HOMOLOGOUS STRUCTURES

Analysis of IPCT homologous structures were addressed extensively by Brito et al., (2011). Briefly, the most similar structures found with Dali server (Holm and Rosenström, 2010) belong to the nucleotidyltransferase family with only 13 to 22% of sequence identity. The first hits are also structures of bifunctional enzymes with highest Z-score of 22 (PDB codes: 4aaw, 2ggq and 1hv9). The N-terminal domain, N-acetylglucosamine-1-phosphate-uridyltransferase, is also a nucleotidyl transferase. However, the second domain, glucosamine-1-phosphate-acetyltransferase, is unrelated to DIPPS and belongs to bacterial transferase hexapeptide family (PF00132).

Superimposition of *A. fulgidus* IPCT structures previously characterized (PDB codes: 2xme and 2xmh, derived from two different crystal forms) with similar structures indicates that the C-terminal helix is missing in the expressed construct, which here is named connector helix 1. The position of this helix appears to be conserved among related structures (Figure 3.12) and forming stable hydrophobic interactions with the core (Figure 3.10a).

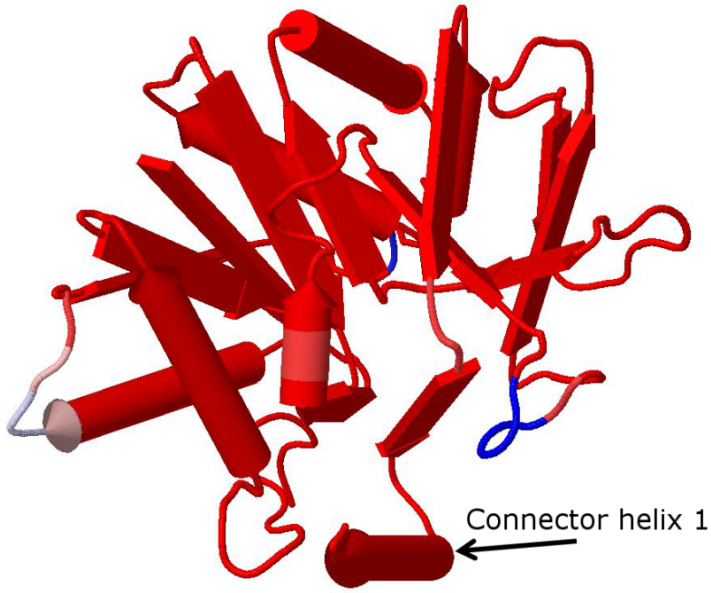


Figure 3.12 Structure conservation of IPCT domain and homologues. IPCT in cartoon rockets representation is colored according to structure conservation (as described by Holm and Rosenström, 2010) within 25 structures similar to IPCT obtained from Dali server (Holm and Rosenström, 2010); red – high structure conservation, blue – low structure conservation. Figure was drawn with Jmol, part of the on-line tool for results analysis in DaliLite. PDB codes of structures included in the comparison: 2ggq, 4aaw, 2ggo, 3pkp, 1hv9, 3hl3, 1g95, 1hm0, 3juj, 3juk, 1hm9, 2oi5, 1hm8, 1g97, 1fzw, 2oi6, 1fxo, 2v0i, 1g1l, 1g23, 1g0r, 1g2v, 1mp5, 2pa4, 1iim, 1mc3.

The absence of connector helix 1 in the previous IPCT structure (Brito et al., 2011) resulted in exposure of the hydrophobic surface of β-sheet formed by β7 and β13. This hydrophobic surface was involved in formation of a strong crystal contact. Similar crystal contact is not present in the full-length IPCT/DIPPS structure. Moreover, due to the strongly hydrophobic character of the connector helix 1 interaction with core of IPCT and conservation of the helix position among

structures of nucleotidyl transferases, the mobility of connector helix 1 and formation of similar interface seems to be unlikely.

The top hits of Dali search for the DIPPS domain was a histidine protein kinase CheA from *Salmonella enterica* (PDB code: 1i5n; rmsd for C^{α} 3.5 Å) and from *Rhodobacter sphaeroides* (PDB code: 3kyj; rmsd for C^{α} 3.9 Å). CheA is histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis (Bell et al., 2010; Mourey et al., 2001). Unlike DIPPS, the CheA is a soluble protein and its α-helices distribution and connectivity resembles that of TM2 – TM6 in DIPPS, despite lack of sequence conservation. The CheA helices corresponding to TM2 – TM5 form a four-helix bundle fold and an additional C-terminal helix is oriented much like TM6 (Figure 3.13).

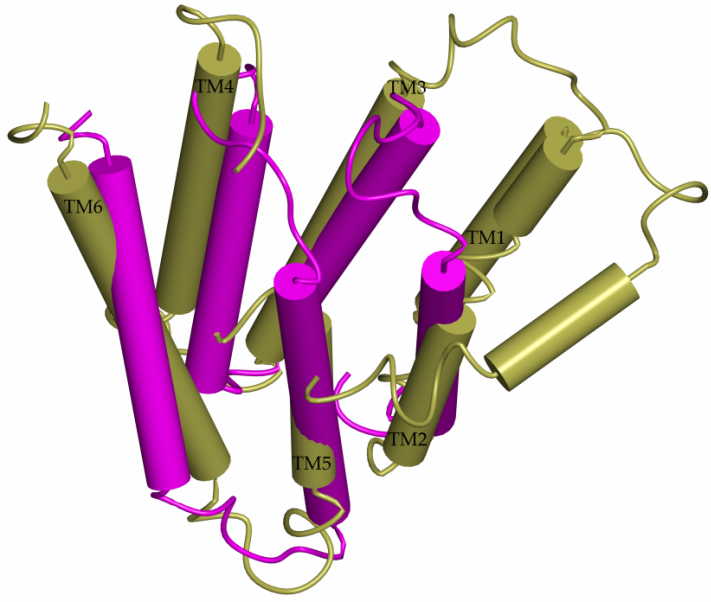


Figure 3.13 Superimposition of DIPPS (gold) with CheA from *Salmonella enterica* (magenta). Helices TM2 - TM5 of DIPPS resemble the four-helix bundle fold of CheA. Helices are shown in cylinder representation. Superimposition performed with Dali server.

3.3.2.7 ACTIVE SITE OF DIPPS

DIPPS catalyses transfer of L-*myo*-inositol from its activated form, CDP-L-*myo*-inositol to another molecule of L-*myo*-inositol-1-phosphate yielding the final product di-*myo*-inositol-1,3´-phosphate-1´-phosphate, CMP and H^+. The DIPPS domain belongs to the CDP-alcohol phosphatidyltransferase family based on the sequence similarity and inclusion of the conserved sequence signature pattern $DG(x)_2AR(x)_8G(x)_3D(x)_3D$ (Figure 3.14, Figure 3.15), starting at Asp360.

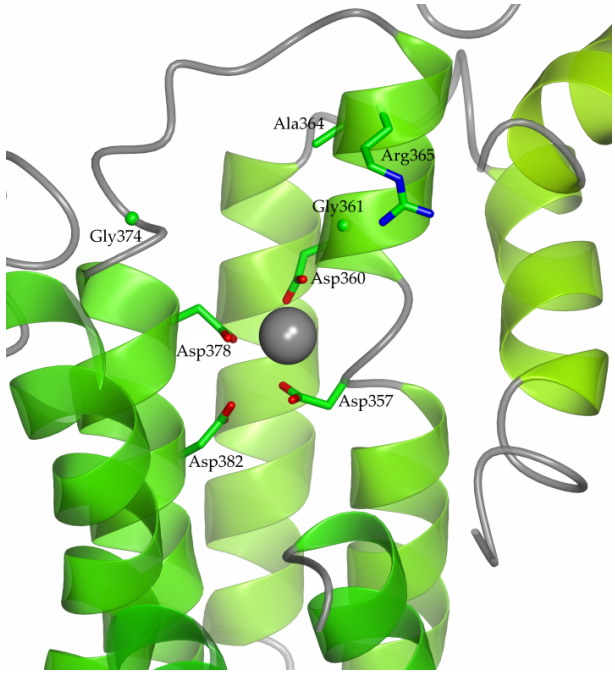


Figure 3.14 Zoomed in view of DIPPS structure shows the amino acid residues from the conserved sequence pattern of the CDP-alcohol phosphatidyltransferase family and conserved Asp357 coordinating the metal ion. The grey sphere indicates putative Mg^{2+}. The atoms are colored as in all the other figures: C – green, O – red, yellow – sulfur, blue – nitrogen, unless indicated otherwise.

This consensus sequence is likely to include some of the catalytic residues crucial in the transfer of the β-phosphoryl from the CDP-alcohol donor. The X-ray structure of IPCT/DIPPS shows a putative magnesium ion site and two of these aspartate residues appear to coordinate Mg^{2+}, with distances of 2.3 Å to Asp360 and 2.6 Å to Asp378. Additionally, the conserved Asp357 (2.4 Å to O^{δ} and 2.3 Å to O of the backbone) is also coordinated to Mg^{2+}. The metal ion was shown to be required for the activity of the enzyme (Table 3.3) and thus should be involved in the catalytic mechanism. Here we assume that the metal ion indicates the position of the active site in the structure (Figure 3.16). The last residue of the consensus sequence, Asp382 is 4.3 Å away from Mg^{2+} ion, at the bottom of the catalytic pocket.

The amino acid residues that are strictly conserved (Figure 3.15) in the enzymes with confirmed or putative DIPPS activity (Gonçalves et al., 2012; Figure 3.16) are likely to play a crucial role in the enzyme function and provide clues into the possible binding sites of the substrates.

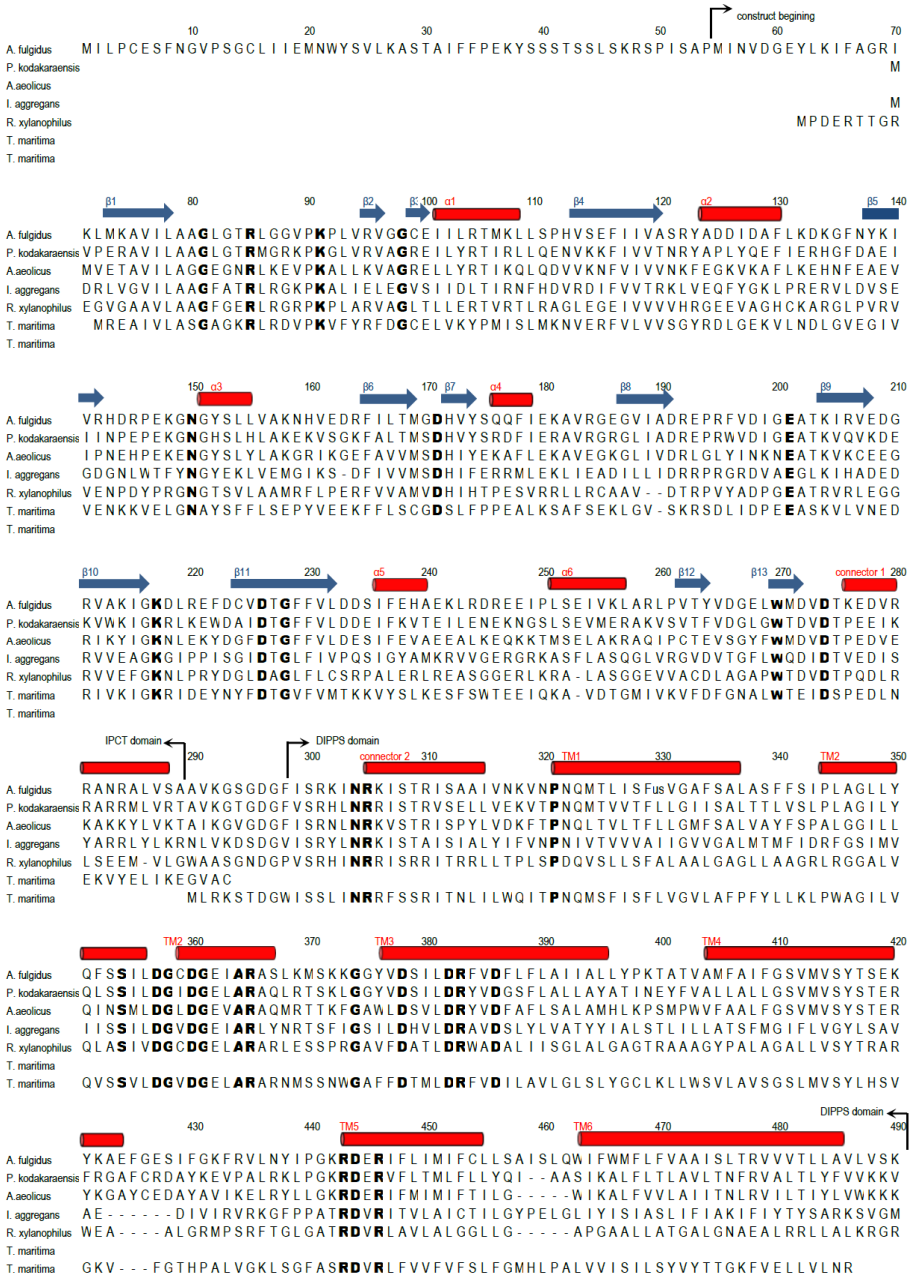


Figure 3.15 Amino acid sequence alignment of IPCT/DIPPS. Created with Geneious alignment (Geneious 6.0.3 software) based on IPCT and DIPPS sequences from 33 organisms with confirmed or putative DIP production (Gonçalves et al., 2012). The sequences represent the major branches of phylogenetic tree (Gonçalves et al., 2012) and the complete alignment is shown in Appendix 1. Black arrows indicate expressed construct beginning and domains

boundaries. Strictly conserved residues among all IPCT/DIPPS sequences (from 33 organisms) are displayed in bold. The secondary structure elements represented with blue arrows - β-strands and red tubes - α-helices were assigned with DSSP (Kabsch and Sander, 1983).

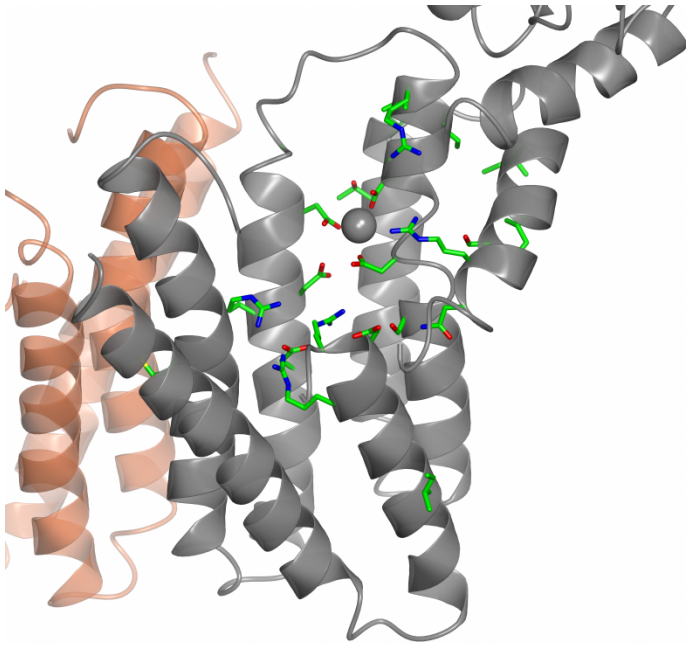


Figure 3.16 DIPPS structure with the conserved residues among the enzymes producing DIP (Figure 3.15) represented as sticks. The atoms are colored: C – green, O – red, S – yellow, N – blue.

It is worth to mention that in the crystal structure the active sites of the IPCT and DIPPS domain do not face each other as one may expect for two enzymes involved in consecutive steps of the synthesis (Figure 3.17). This could be an effect of the crystal packing and it is also not clear if there could be some rearrangement within the enzyme that will allow direct passage of the product of the IPCT domain into the active site of the DIPPS domain.

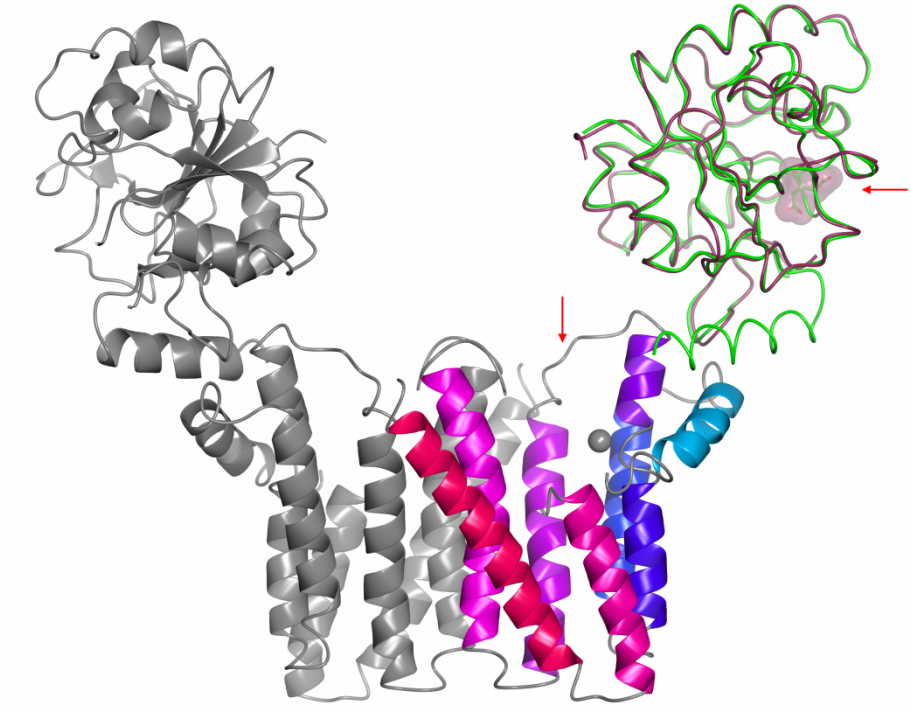


Figure 3.17 Arrows indicate the accessibility towards the active sites of IPCT and DIPPS. The IPCT domain is in ribbon representation (green) and superimposed with the structure of IPCT alone previously determined and shown in dark purple (2XMH, Brito et al., 2011) with citrate bound in active site (indicated with transparent purple surface).

3.3.2.8 PUTATIVE SUBSTRATE BINDING SITE

The binding site of the substrate, L-*myo*-inositol 1-phosphate, in the DIPPS domain was probed with HEX docking server (Macindoe et al., 2010). The plausible solution was chosen based on the ligand fitting, its interactions and position with respect to the conserved residues and putative catalytic site (Figure 3.18).

The modeled L-*myo*-inositol 1-phosphate accommodates the pocket at the bottom of the active site. The phosphoryl group of the substrate interacts with the conserved positively charged Arg305, and conserved Ser308 is 4.3 Å away from the phosphoryl group. Nearby, the conserved Gly358 introduces a kink into TM2 but also enables formation of the cavity to accommodate the substrate´s phosphoryl. The inositol ring forms hydrogen bonds with the strictly conserved Asn304, Arg443, Ser354 and Asp357. The hydroxyl group located on the C3 of the inositol ring, which is engaged in the reaction, forms H-bond to conserved Asp357. In addition, Asp382 of the consensus sequence is 5.5 Å away from the hydroxyl.

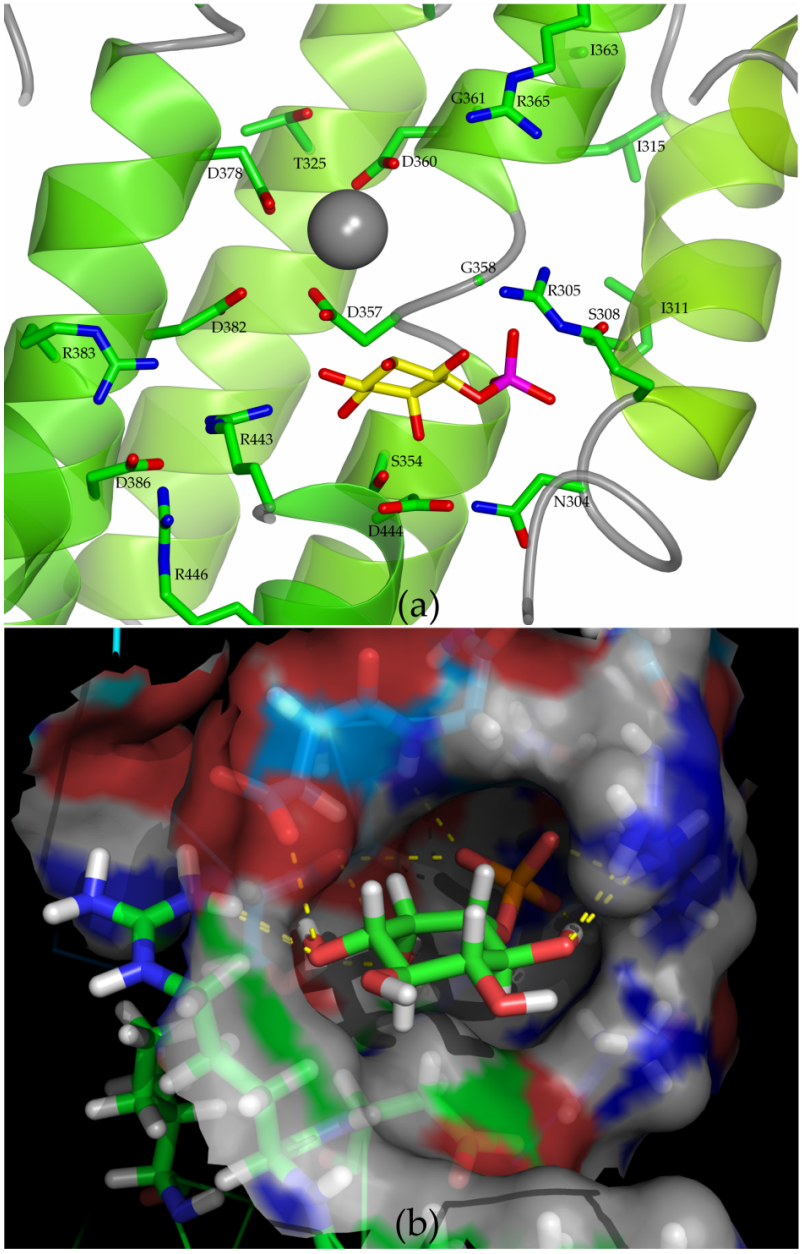


Figure 3.18 L-*myo*-inositol 1-phosphate docked into DIPPS represented as (a) cartoon representing α-helices and side chains of strictly conserved residues in sticks mode (b) van der Waals surface of the binding pocket. In the figure (a) the C atoms of the substrate are colored in yellow. The other atoms are colored: C – green, O – red, S – yellow, N – blue, H – white, P – magenta in (a) and orange in (b). Figure (a) was drawn with CCP4MG (McNicholas et al., 2011) and (b) with PyMOL.

3.3.2.9 MUTAGENESIS

The structure of IPCT/DIPPS provided the first atomic model for CDP-alcohol phosphatidyltransferase family. This model was used in an attempt to understand the role of some of the residues which might be involved in the catalytic mechanism of the enzyme. Single point mutations have been introduced into DIPPS domain as listed in Table 3.3 and their impact on the activity was measured. Detailed results of the activity tests can be found in Appendix 2.

Table 3.3 Impact of the IPCT/DIPPS point mutations on the DIPPS activity. CSP – consensus sequence pattern of the CDP-alcohol phosphatidyltransferase family. Activity: moderate reduction – 80% - 30% of the WT activity, strong reduction – below 30% of the WT activity. Detailed results are presented in the Appendix 2.

mutation	activity	CSP	localization
S308A	moderate reduction		connector helix 2
S308M	inactivation		connector helix 2
D357A	inactivation		TM2
D357N	strong reduction		TM2
D360A	inactivation	yes	TM2
D360N	inactivation	yes	TM2
A364M	strong reduction	yes	TM2
R365M	strong reduction	yes	TM2
S371P	strong reduction		Loop TM2-TM3
D378A	inactivation	yes	TM3
D378N	inactivation	yes	TM3
D382A	inactivation	yes	TM3
D382N	inactivation	yes	TM3
R443A	inactivation		TM5
WT(-Mg^{2+})	inactivation		

The residues tested by mutagenesis can be divided into three groups with respect to their sequence conservation. The group comprising the conserved consensus sequence pattern (CSP) of the family (Table 3.3) may be the most critical for the reaction mechanism, independently from the substrate utilized by the enzyme. The second group of residues consists of residues strictly conserved within the sequences of known or putative DIPPS function as described by Gonçalves et al., 2012 (Figure 3.15). Some of these may be involved in the substrate binding. The third group of residues that are not strictly conserved and so their function may not be limited to the exact side chain properties and their role might be fulfilled by other similar (or not) residues. The positions of mutated residues in the structure are shown in the Figure 3.19. In further discussion, the conserved residues refer to the enzymes with confirmed or putative DIPPS activity (Gonçalves et al., 2012). The residues conserved in the CDP-alcohol phosphatidyltransferase family are referred as consensus sequence pattern or CSP.

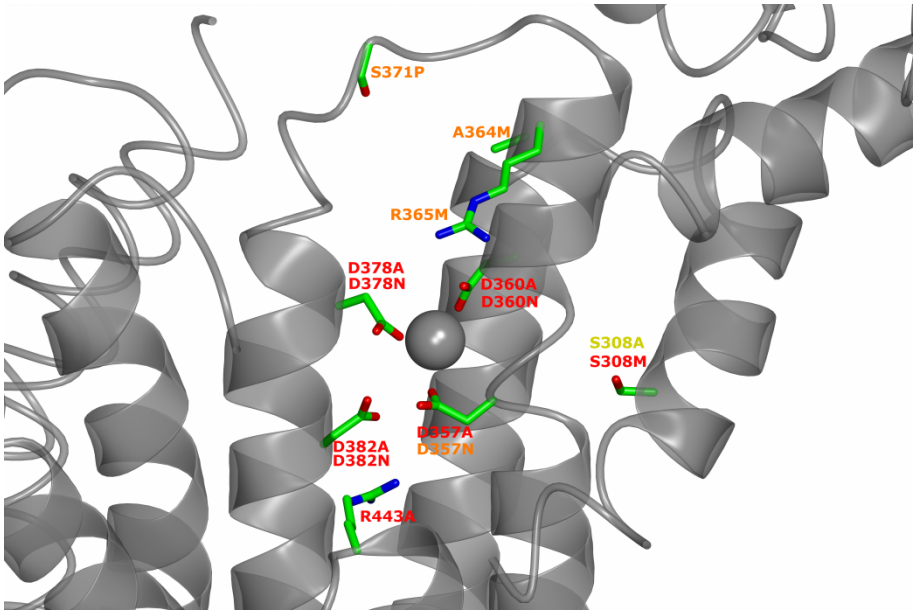


Figure 3.19 Residues mutated for activity assay are shown as sticks. The color of the label corresponds to the effect on DIPPS activity; yellow – moderate activity reduction, orange - strong reduction and red – inactivation. The atoms are colored: C – green, O – red, N – blue.

All the conserved aspartate residues located nearby the putative Mg^{2+} site, when mutated to alanine, inactivate the DIPPS enzyme (Figure 3.19). However, all mutations of those residues to asparagines inactivate the enzyme, except Asp357, which retains some basal activity. Effect of Asp357, Asp360 and Asp378 mutations on the activity might be related to their role in coordination to the Mg^{2+} ion. This metal ion was shown to be required for reaction and the activity assays were performed in the presence of 10 mM $MgCl_2$, while in the absence of metal ion the enzyme is inactive. Defective binding of Mg^{2+} might be a plausible explanation for the observed inactivation of D357A, D360A/N, and D378A/N or severe decrease of activity of D357N.

The role of the Asp357 is likely to be beyond anchoring the Mg^{2+} ion. In fact this residue is not included into the CSP. However, it is well conserved in related protein sequences, including the five human enzymes belonging to CDP-alcohol phosphatidyltransferase family. As earlier discussed, this residue interacts with the docked L-*myo*-inositol 1-phosphate. The H-bonding is to the hydroxyl located on C3 of inositol ring, which is involved in the reaction of coupling with another L-*myo*-inositol 1-phosphate molecule activated with CDP. Preservation of basal activity for one of the mutants, D357N, suggests its role in positioning the substrate for reaction, which can still be partially maintained by asparagine, but abolished with alanine residue.

The aspartate Asp382 (CSP) is located 4.2 Å away from Mg^{2+} and 5.5 Å away from the docked L-*myo*-inositol 1-phosphate (Figure 3.18). Mutation of Asp382 to alanine or asparagine, inactivating DIPPS, suggests that Asp382 role might be not just positioning through H-bonding of substrate, which might be still carried out by asparagine. At this initial stage it is assumed that Asp382 may perform some catalytic function.

The binding mode of the L-*myo*-inositol 1-phosphate as obtained from docking (Figure 3.18) was probed with mutation of Ser308 of connector helix 1, located at the end of the putative substrate binding pocket. Position covered from the putative catalytic site by TM2 suggests that the residue is not directly involved in catalysis. In our model this residue may be involved in the H-bonding to the

phosphoryl of L-*myo*-inositol 1-phosphate. Mutation S308A results in moderate reduction of DIPPS activity. We speculate that this may be a result of weakening of L-*myo*-inositol 1-phosphate binding due to lost of the H-bond between the substrate and Ser308 in the binding site. Introducing longer methionine residue, S308M, inactivates the enzyme possibly by partially filling in the phosphoryl pocket binding and thus interfering in L-*myo*-inositol 1-phosphate binding. Additionally, Ser308 is a strictly conserved residue in enzymes with DIPPS activity and thus it may well be important for interaction with substrate.

The second serine residue tested is Ser371 located in the loop connecting TM2 and TM3 (residues 367 - 376). This loop contains Gly374 which is part of the consensus sequence pattern. In this particular enzyme there is also a second neighboring glycine, Gly375. Comparison with other sequences of the family shows that it is common to find additional glycine residues in this loop. Strict conservation of the Gly374 in the enzyme family may indicate the importance of flexibility of this loop for function. For mutagenesis we have chosen Ser371, which is not strictly conserved and may appear as threonine or glutamine in the DIPPS enzymes or other residues in enzymes of the family. This serine, located in the middle of the loop and mutated to proline was expected to affect loop flexibility. In result of the S371P mutation, we observed strong decrease of the activity. Similar mutation was described in an allele of another member of the family, phosphatidylglycerophosphate synthase from

E. coli (Usui et al., 1994). The corresponding Thr60 was also mutated to proline resulting in lower enzyme activity and causing acidic-phospholipid deficiency.

Interestingly, the loop (interconnecting TM2-TM3) comprises three lysine residues Lys269, Lys372 and Lys373. Positively charged residues and flexibility may suggest a role of the loop in attracting the negatively charged substrates from solution and maybe the initial binding of substrates before passing to the deeper location of the active site binding pocket. In addition to the positively charged loop at the entrance to the putative active site, there is also Arg365, part of the conserved sequence pattern. Similarly, its role might involve attracting substrates from solution and positioning the substrate in the active site. We found that its positive charge is important for the activity and its mutation to methionine resulted in strong reduction of activity.

Ala364 is also part of the CSP and is located next to Arg365, and roughly beneath the loop between TM2 and TM3. Its short side chain may enable more mobility for the loop or even provide space useful for transient binding of the substrate by the flexible loop, and hence mutant A364M was found to strongly reduce the enzyme activity.

The bottom of the hydrophilic cavity inside the DIPPS domain is formed by three strictly conserved arginine residues, Arg383, Arg443 and Arg446. The last two are anchored in place by the strictly conserved Asp386, which indicate that positions of those arginines may be important. In fact, mutation of R443A inactivates the enzyme.

Arg443 is only 4.4 Å away from Asp382 (important for catalysis), and 3.2 Å away from Asp357 (H-bonding to hydroxyl on C3 of docked L-*myo*-inositol 1-phosphate, which is involved in reaction). It is proposed that Arg443 provides the binding site for the β phosphate of CDP-L-*myo*-inositol, which is involved in the reaction to give the final adduct with L-*myo*-inositol 1-phosphate.

Mutagenesis of residues from the CSP to alanine or glycine was previously described for *Saccharomyces cerevisiae* cholinephosphotransferase (CPT1) (Williams and McMaster, 1998). The striking difference in the obtained results concerns the first residue of the CSP, the counterpart of Asp360 (residue D113 of CPT1). In DIPPS, both mutations to alanine and asparagine inactivate the enzyme, while in CPT1 has almost no effect on its activity. This could indicate that Asp360 has a more specific role for the type of substrate. An alternative explanation could be that the high temperature (85°C) at which the activity tests are performed may require stronger binding of Mg^{2+} in order to keep it in the active site for reaction. Mutations to alanine of the other two aspartates corresponding to Asp378, Asp382 (Asp131, Asp135 of CPT1) inactivate both enzymes.

DIPPS mutations A364M, R365M and mutations of corresponding residues of CPT1 (Ala117, Arg118) to glycine and alanine respectively, reduced strongly the activity in all the cases. While mutation of Ala364 to methionine imposes additional steric hindrance, its mutation to glycine must influence the reaction in a

different way. Possibly, by destabilizing the secondary structure at the top of the TM2 and thus displacing the Arg365.

In our study the conserved glycine residues were not tested, however very interesting results were obtained for CPT1. The counterpart of Gly361 (Gly114 in CPT1) mutated to alanine abolished enzyme activity. This glycine is positioned next to the first aspartate of the consensus pattern and one helix turn above is located the arginine of consensus pattern. Alanine side chain introduces small additional space restriction and also limits the flexibility of the helix backbone. The inactivating effect of this mutation may indicate that in the course of reaction there is some significant conformational change required, which is dependent on the presence of the conserved glycine.

Mutation to alanine of another glycine (Gly127 in CPT1, corresponding to Gly374 of DIPPS) of the CSP also abolished activity underlining the crucial role of the flexible loop between TM2 and TM3.

Detailed analysis of the enzyme mechanism and function of particular residues is at this stage impaired by lack of experimental information on the substrates binding mode. It was possible to dock L-*myo*-inositol 1-phosphate (Figure 3.18) into the well conserved pocket of DIPPS domain. However, reasonable positioning of the second substrate appears to be more troublesome.

The CDP-alcohol phosphatidyltransferase family of enzymes perform the same type of reaction thus may utilize common catalytic residues.

Moreover, one of the substrates is activated for reaction by coupling with CDP. This implies that some of the family consensus pattern residues may be involved in forming the binding site of the CDP fragment.

In the first approach it would be reasonable to assume binding of the CDP fragment to a site within the loop between TM2 and TM3, and upper part of those helices where the CSP residues are located. However, the apparent flexibility of the loop may result in adopting the proper conformation of polypeptide for substrate binding only in the presence of the substrate.

If that is indeed the case and CDP binds in the loop, then possibly the β phosphate of CDP could coordinate to Mg^{2+}, to be activated for the nucleophilic attack by the hydroxyl of the L-*myo*-inositol 1-phosphate, as the likely mechanism relies on a single displacement reaction (Raetz et al., 1987). That will leave relatively large opening in the center of DIPPS domain just for inositol fragment. Thus an alternative binding position of the CDP fragment of substrate is in the bottom of the active site opening, which is formed by three conserved arginine residues (one of them mutated to alanine inactivated DIPPS).

CONCLUSIONS

Herein, we present the first structure of CDP-alcohol phosphatidyltransferase family representative, CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (DIPPS). It is fused with L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) domain, which activates the L-*myo*-inositol 1-phosphate for the second phosphotransfer reaction. The IPCT domain displays the typical features of the nucleotidyltransferases with extended Rossmann fold pattern. The DIPPS domain is integral to the membrane, with 6 transmembrane helices and one amphipathic helix positioned approximately parallel to the membrane surface. The helices form approximately 20 Å deep hydrophilic opening in the domain, which contains residues of the conserved sequence pattern of CDP-alcohol phosphatidyltransferase family and putative Mg^{2+} binding site. The dimeric quaternary arrangement is formed through an extensive interface involving three of the transmembrane helices. The IPCT binds weakly to the solvent exposed edge of DIPPS domain in an arrangement in which the two active sites do not face each other.

In an attempt to shed light on the location and the role of active site residues, L-*myo*-inositol 1-phosphate was docked into the DIPPS domain and mutagenesis studies allowed identification of several residues important for DIPPS activity.

DIPPS is a member of the CDP-alcohol phosphatidyltransferase family, for which no structure was yet available. This family

comprises integral membrane proteins, such as choline/ethanolamine phosphotransferase and cardiolipin synthase. Members of this family are present in all domains of life and usually involved in phospholipid biosynthesis. The X-ray structure of IPCT/DIPPS determined at 2.8 Å resolution will have a significant impact in the field of lipid biosynthesis and will pave the way to further studies, for example, exploring the mechanism of reaction, homology modeling followed by structure-based drug design.

ACKNOWLEDGEMENT

The authors would like to thank to João Carita for large scale fermentations and Danielle de Sanctis for helping with crystal diffraction screening. The work was supported by EU grant (Marie Curie, FP7/2007-2013, no. 211800), by Fundação para a Ciência e a Tecnologia (FCT) through grants PTDC/BIA-MIC/099963/2008 and PTDC/BIA-PRO/103718/2008. We acknowledge Diamond, ESRF and SLS synchrotron for financial support for data collection.

REFERENCES

Bell, C.H., Porter, S.L., Strawson, A., Stuart, D.I., Armitage, J.P., 2010. Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. PLoS Biol. 8, e1000306.

Berry, I.M., Dym, O., Esnouf, R.M., Harlos, K., Meged, R., Perrakis, A., Sussman, J.L., Walter, T.S., Wilson, J., Messerschmidt, A., 2006. SPINE high-throughput crystallization, crystal imaging and recognition techniques: current state, performance analysis, new technologies and future aspects. Acta Crystallogr. D Biol. Crystallogr. 62, 1137–1149.

Bill, R.M., Henderson, P.J.F., Iwata, S., Kunji, E.R.S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G., Vogel, H., 2011. Overcoming barriers to membrane protein structure determination. Nat. Biotechnol. 29, 335–340.

Borges, N., Gonçalves, L.G., Rodrigues, M.V., Siopa, F., Ventura, R., Maycock, C., Lamosa, P., Santos, H., 2006. Biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile Archaeoglobus fulgidus in stress adaptation. J. Bacteriol. 188, 8128–8135.

Brito, J.A., Borges, N., Vonrhein, C., Santos, H., Archer, M., 2011. Crystal structure of Archaeoglobus fulgidus CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper)thermophiles. J. Bacteriol. 193, 2177–2185.

Empadinhas, N., Da Costa, M.S., 2006. Diversity and biosynthesis of compatible solutes in hyper/thermophiles. Int. Microbiol. 9, 199–206.

Evans, P.R., 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D 67, 282–292.

Geerlof, A., Brown, J., Coutard, B., Egloff, M.P., Enguita, F.J., Fogg, M.J., Gilbert, R.J.C., Groves, M.R., Haouz, A., Nettleship, J.E., Nordlund, P., Owens, R.J., Ruff, M., Sainsbury, S., Svergun, D.I., Wilmanns, M., 2006. The impact of protein characterization in structural proteomics. Acta Crystallogr. D Biol. Crystallogr. 62, 1125–1136.

Gonçalves, L.G., Borges, N., Serra, F., Fernandes, P.L., Dopazo, H., Santos, H., 2012. Evolution of the biosynthesis of di-myo-inositol phosphate, a marker of adaptation to hot marine environments. Environ. Microbiol. 14, 691–701.

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–549.

Kabsch, W., 2010. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Kamtekar, S., Hecht, M.H., 1995. Protein Motifs. 7. The four-helix bundle: what determines a fold? FASEB J. 9, 1013–1022.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Kunji, E.R.S., Harding, M., Butler, P.J.G., Akamine, P., 2008. Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. Methods 46, 62–72.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 40, D370–376.

Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M.-D., Ritchie, D.W., 2010. HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res. 38, W445–449.

Mascellani, N., Liu, X., Rossi, S., Marchesini, J., Valentini, D., Arcelli, D., Taccioli, C., Helmer Citterich, M., Liu, C.-G., Evangelisti, R., Russo, G., Santos, J.M., Croce, C.M., Volinia, S., 2007. Compatible solutes from hyperthermophiles improve the quality of DNA microarrays. BMC Biotechnol. 7, 82.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr. D Biol. Crystallogr. 67, 386–394.

Mourey, L., Da Re, S., Pédelacq, J.D., Tolstykh, T., Faurie, C., Guillet, V., Stock, J.B., Samama, J.P., 2001. Crystal structure of the CheA

histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. J. Biol. Chem. 276, 31074–31082.

Ostermeier, C., Michel, H., 1997. Crystallization of membrane proteins. Curr. Opin. Struct. Biol. 7, 697–701.

Proteau, A., Shi, R., Cygler, M., 2010. Application of dynamic light scattering in protein crystallization. Curr Protoc Protein Sci Chapter 17, Unit 17.10.

Raetz, C.R., Carman, G.M., Dowhan, W., Jiang, R.T., Waszkuc, W., Loffredo, W., Tsai, M.D., 1987. Phospholipids chiral at phosphorus. Steric course of the reactions catalyzed by phosphatidylserine synthase from Escherichia coli and yeast. Biochemistry 26, 4022–4027.

Rao, S.T., Rossmann, M.G., 1973. Comparison of super-secondary structures in proteins. J. Mol. Biol. 76, 241–256.

Rodionov, D.A., Kurnasov, O.V., Stec, B., Wang, Y., Roberts, M.F., Osterman, A.L., 2007. Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate. Proc. Natl. Acad. Sci. U.S.A. 104, 4279–4284.

Rodrigues, M.V., Borges, N., Henriques, M., Lamosa, P., Ventura, R., Fernandes, C., Empadinhas, N., Maycock, C., Da Costa, M.S., Santos, H., 2007. Bifunctional CTP:inositol-1-phosphate cytidylyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles. J. Bacteriol. 189, 5405–5412.

Santos, H., Lamosa, P., Borges, N., Gonçalves, L.G., Pais, T., Rodrigues, M.V., 2011. Organic Compatible Solutes of Prokaryotes that Thrive in Hot Environments: The Importance of Ionic Compounds for Thermostabilization, in: Horikoshi, K. (Ed.), Extremophiles Handbook. Springer Japan, Tokyo, pp. 497–520.

Siddiqui, K.S., Thomas, T., 2008. Protein adaptation in extremophiles. Nova Biomedical Books, New York.

Stetter, K.O., 1999. Extremophiles and their adaptation to hot environments. FEBS Lett. 452, 22–25.

Stroud, R.M., 2011. New tools in membrane protein determination. F1000 Biol Rep 3, 8.

Trivedi, S., Gehlot, H.S., Rao, S.R., 2006. Protein thermostability in Archaea and Eubacteria. Genet. Mol. Res. 5, 816–827.

Usui, M., Sembongi, H., Matsuzaki, H., Matsumoto, K., Shibuya, I., 1994. Primary structures of the wild-type and mutant alleles encoding the phosphatidylglycerophosphate synthase of Escherichia coli. J. Bacteriol. 176, 3389–3392.

Williams, J.G., McMaster, C.R., 1998. Scanning alanine mutagenesis of the CDP-alcohol phosphotransferase motif of Saccharomyces cerevisiae cholinephosphotransferase. J. Biol. Chem. 273, 13482–13487.

Concluding remarks

This thesis describes the protein production and structures of two new enzymes; a bifunctional CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) and CDP-L-*myo*-inositol *myo*-inositolphosphotransferase (DIPPS) from *Archaeoglobus fulgidus* at 2.8 Å resolution, and an α-phosphoglucomutase from *Lactococcus lactis* at 1.5 Å resolution.

The IPCT/DIPPS was a challenging target to study as its DIPPS domain is integral to the membrane. Membrane embedded hydrophobic helices contribute to a number of obstacles in protein production and crystallization, which explains the very limited number of membrane protein structures available to date. In particular, the crystal diffraction quality varies significantly within the same crystallization conditions and thus numerous crystals need to be screened for diffraction, with considerable synchrotron beamtime required. In a collaborative effort these problems were overcome and the first structural insights into the CDP-alcohol phosphatidyltransferase family have been achieved. The novel DIPPS structure contains six transmembrane helices, unlike all the attempted sequence-based transmembrane region predictions (2 - 5 TMs). It reveals a hydrophilic cavity with active site residues in the central part of the domain and opens to the cytoplasm. Crucial residues for the enzyme activity have been identified by mutagenesis studies and docking of L-*myo*-inositol 1-phosphate provided preliminary insights into the substrate binding mode. Further studies will follow to provide structural information on the enzyme

complexes with its substrates. That in turn will facilitate exciting studies involving mutagenesis and computer modeling aiming at the detailed description of reaction mechanism with impact on the fields of compatible solute and lipid biosynthesis. Interesting new opportunity also opens for computational modeling studies, which now can provide homology models for some relevant related proteins.

A study that can be facilitated by the available structure concerns IPCT/DIPPS activation at temperatures over 60°C, with maximum activity at supraoptimal temperatures for (hyper)thermophiles. The related lipid-biosynthesizing enzymes from mesophilic organisms do not require elevated temperatures for catalysis of the same type of reaction. This points to a possibility of temperature induced mechanism for enzyme activation, which might be probed in a search of structural changes e.g. by NMR.

Another open question concerns the possible cooperation between the two domains with consecutive activities, IPCT and DIPPS, of which active sites in the obtained structure face apart from each other.

It is worth mentioning that the majority of CDP-alcohol phosphatidyltransferase family members comprise a single integral membrane domain. This might have been a disadvantage in the crystallization trials thus far. In this study, the bifunctional IPCT/DIPPS construct proved to be crucial in the crystallization as the crystal contacts were formed mainly through the cytoplasmic IPCT

domain. This result may indicate the direction worth exploring in structural studies of other members of CDP-alcohol phosphatidyltransferase family.

The novelty of the second structure described in this thesis, α-phosphoglucomutase from *Lactococcus lactis* (APGM), is in its substrate specificity. It is the first structurally characterized bacterial α-phosphoglucomutase with structure unlike the common α-phosphoglucomutases of α-D-phosphohexomutase superfamily. Based on the sequence similarity, conservation of the catalytic residues and in particular the structural similarity, APGM may be regarded as eukaryotic phosphomannomutase (PMM), which lost its ability to utilize mannose 1-phosphate. In this context the modeled α-glucose 1-phosphate in the active site compared with the structures of PMMs provided very interesting clues regarding the substrate discrimination. This lead to a proposal that single residue changes at the otherwise conserved substrate binding site may be responsible for the differentiation between the two epimeric compounds, mannose 1-phosphate and α-glucose 1-phosphate of APGM. The obvious next step should include mutagenesis study to assay if with single mutations the APGM activity towards mannose 1-phosphate could be observed.

Other interesting experiments that could be performed concern the quaternary structure of APGM, which in the crystal structure appears to form a novel dimeric assembly as compared to other eukaryotic

PMMs. Mutagenesis on the putative dimeric interface could be used in an attempt to disrupt the observed dimer in solution. Moreover, upon dimer formation in the center between the two active sites a patch of positive charge is observed and mutagenesis could give information if this charge provides some advantage with regards to the enzymatic efficiency.

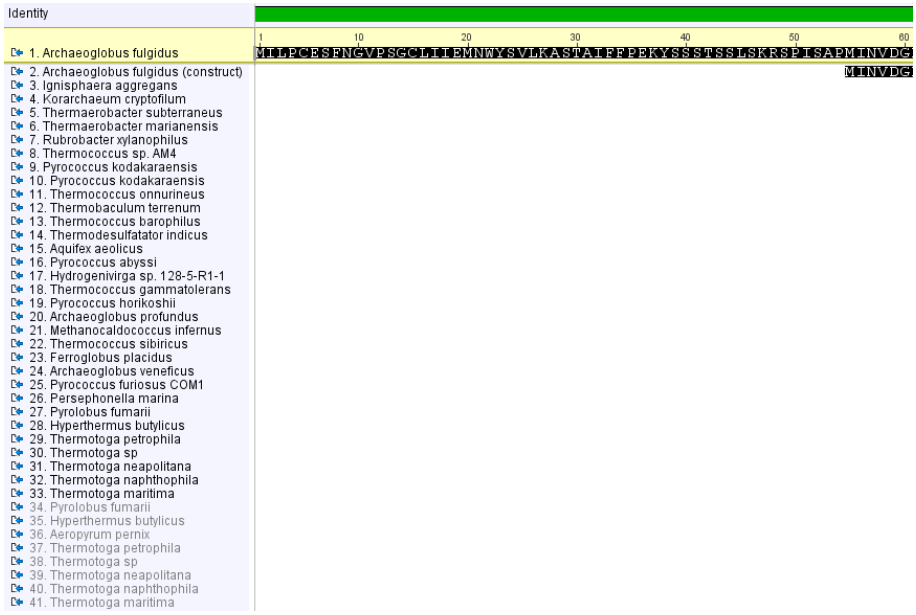
X-ray crystallography provides invaluable information on the atomic details of biological macromolecules with two new structures being described here. It is the structural biologist task to interpret those protein models and find clues regarding their relevance for the protein function. This task has been developed in this work and resulted in a number of new questions, which still need to be addressed with further crystallographic and biochemical studies. The efforts to better understand the structure-function relation of proteins described here will be continued in our Laboratory, so one can anticipate further exciting results emerging from the work developed in the course of this thesis.

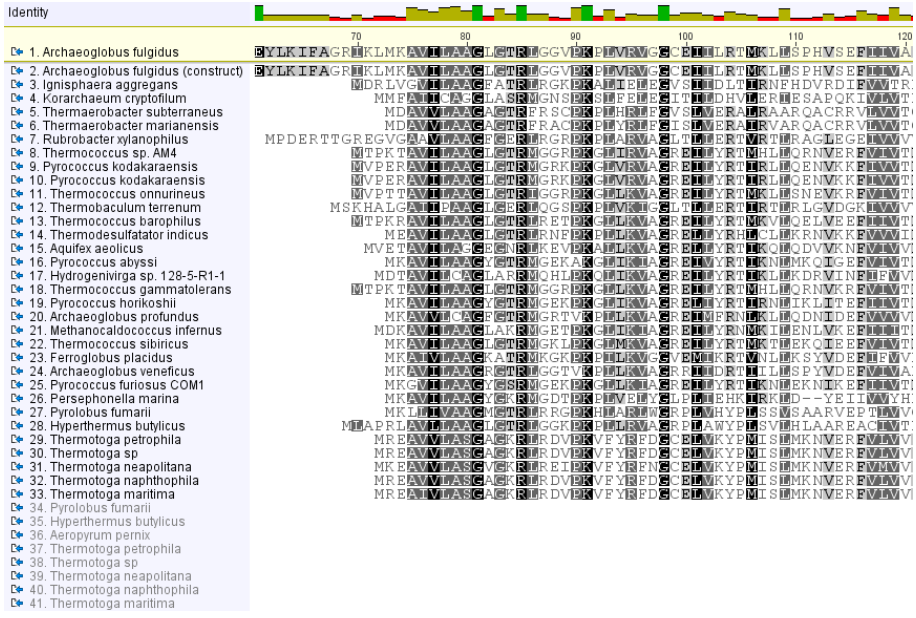
Appendices

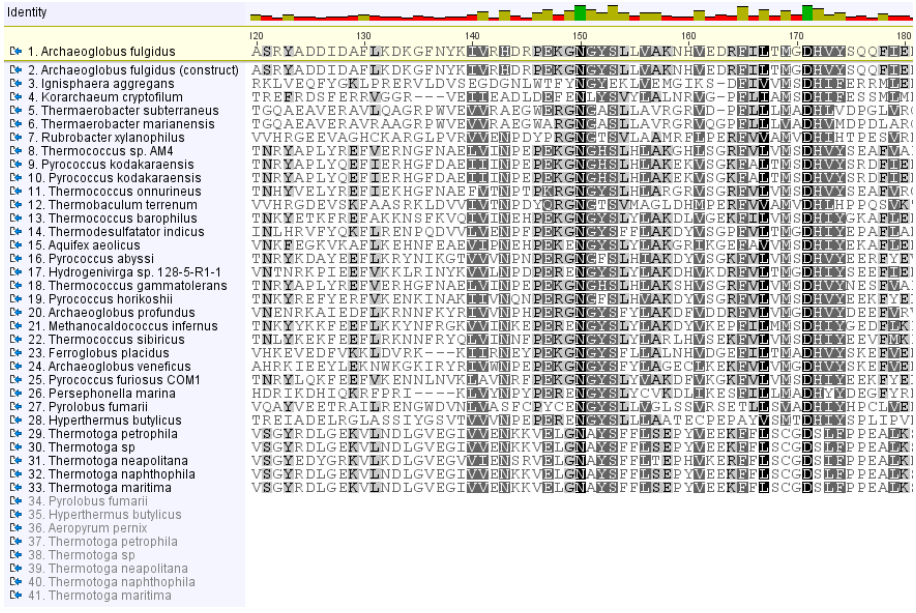
Appendix 1 Sequence alignment of proteins with IPCT/DIPPS activity.

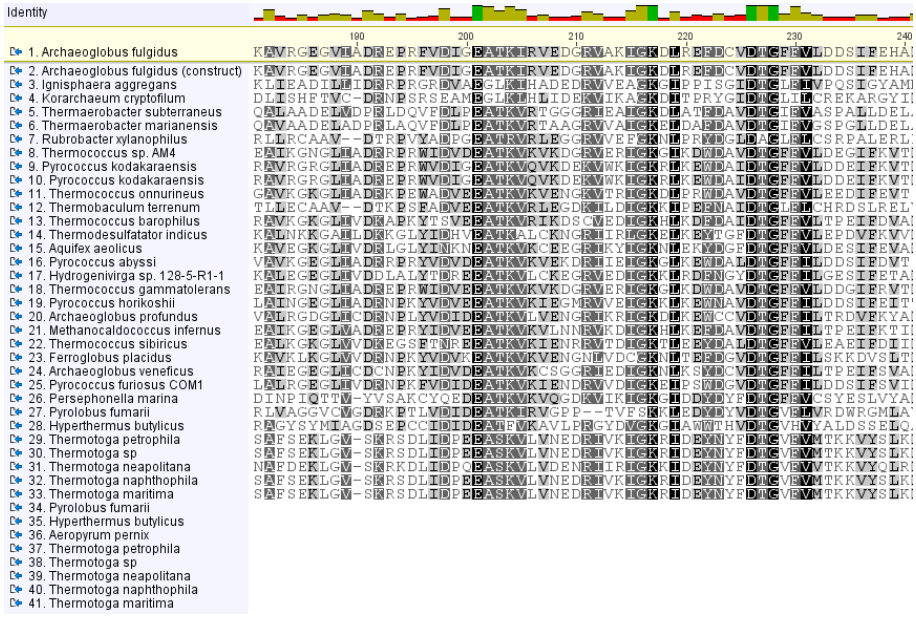
Created with Geneious alignment (Geneious 6.0.3 software) based on IPCT and DIPPS sequences from 33 organisms with confirmed or putative DIP production (Gonçalves et al., 2012). The colored bar above alignment indicates conservation of the residues and from green for strict conservation and red no conservation. Additionally, the residue symbols are in black boxes for strictly conserved to grey for similar residues. The 1st sequence is a gene annotated as IPCT/DIPPS of *Archaeoglobus fulgidus* and the 2nd sequence is the construct described in this study and expressed with additional tags for purification.

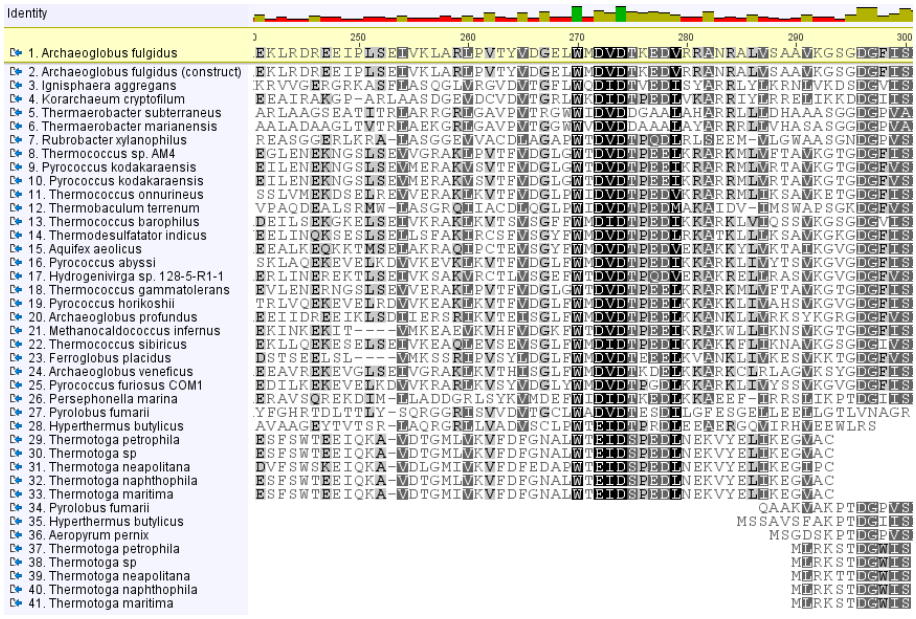
UniProt codes of the sequences: O29976_ARCFU (*Archaeoglobus fulgidus*), E0SRX5_IGNAA (*Ignisphaera aggregans*), B1L3B5_KORCO (*Korarchaeum cryptofilum*), E4M3F3_9FIRM (*Thermaerobacter subterraneus)*, E6SL49_THEM7 (*Thermaerobacter marianensis*), Q1AWQ0_RUBXD (*Rubrobacter xylanophilus*), B7R372_9EURY (*Thermococcus sp. AM4*), Q5JDA9_PYRKO (*Pyrococcus kodakaraensis*), B6YXC3_THEON (*Thermococcus onnurineus*), D1CEN7_THET1 (*Thermobaculum terrenum*), F0LMX7_THEBM (Thermococcus barophilus), F8A8W4_THEID (Thermodesulfatator indicus), O67379_AQUAE (*Aquifex aeolicus*), Q9UZW1_PYRAB (*Pyrococcus abyssi*), A8UUK0_9AQUI (*Hydrogenivirga sp. 128-5-R1-1*), C5A1T6_THEGJ (*Thermococcus gammatolerans*), O58964_PYRHO (*Pyrococcus horikoshii*), D2RHQ4_ARCPA (*Archaeoglobus profundus*), D5VRC1_METIM (*Methanocaldococcus infernus*), C6A371_THESM (*Thermococcus sibiricus*), D3S199_FERPA (*Ferroglobus placidus*), F2KS67_ARCVS (*Archaeoglobus veneficus*), I6TWS7_9EURY (*Pyrococcus furiosus COM1*), C0QSN6_PERMH (*Persephonella marina*), G0EE02_PYRF1 G0EE03_PYRF1 (*Pyrolobus fumarii*), Q9YBT3_AERPE (*Aeropyrum pernix*), A2BJ61_HYPBU A2BJ62_HYPBU (*Hyperthermus butylicus*), A5IMG5_THEP1 A5IMG6_THEP1 (*Thermotoga petrophila*), B1LBF8_THESQ B1LBF7_THESQ (*Thermotoga sp*), B9K8G7_THENN B9K8G8_THENN (*Thermotoga neapolitana*), D2C420_THENR D2C421_THENR (*Thermotoga naphthophila*), G4FFF3 G4FFF4 (*Thermotoga maritime*).

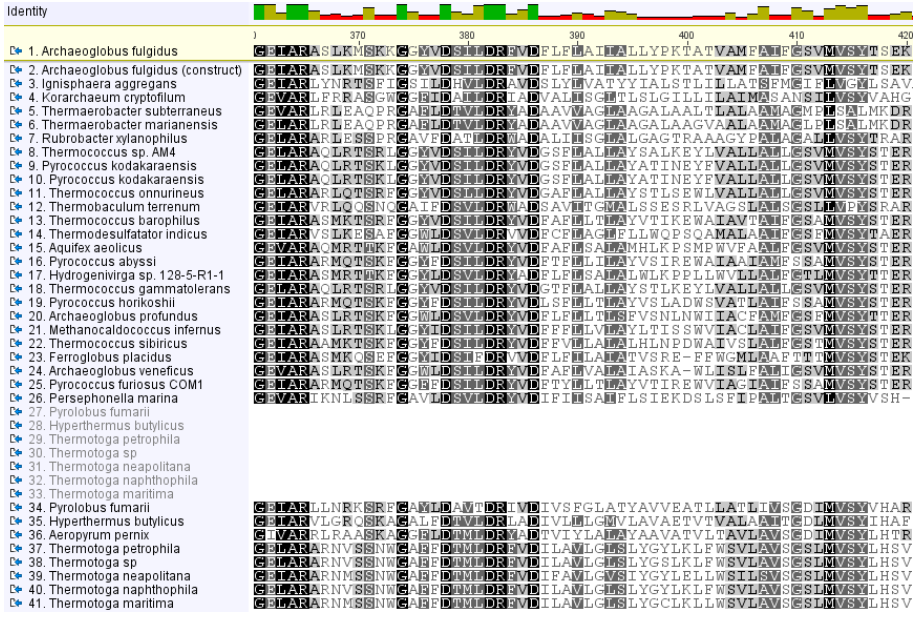

Identity
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima


Identity
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima


Identity
120
130
140
150
160
170
180
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima


Identity
190
200
210
220
230
240
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima


Identity
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima


Identity
370
380
390
400
410
420
1. Archaeoglobus fulgidus
2. Archaeoglobus fulgidus (construct)
3. Ignisphaera aggregans
4. Korarchaeum cryptofilum
5. Thermaerobacter subterraneus
6. Thermaerobacter marianensis
7. Rubrobacter xylanophilus
8. Thermococcus sp. AM4
9. Pyrococcus kodakaraensis
10. Pyrococcus kodakaraensis
11. Thermococcus onnurineus
12. Thermobaculum terrenum
13. Thermococcus barophilus
14. Thermodesulfatator indicus
15. Aquifex aeolicus
16. Pyrococcus abyssi
17. Hydrogenivirga sp. 128-5-R1-1
18. Thermococcus gammatolerans
19. Pyrococcus horikoshii
20. Archaeoglobus profundus
21. Methanocaldococcus infernus
22. Thermococcus sibiricus
23. Ferroglobus placidus
24. Archaeoglobus veneficus
25. Pyrococcus furiosus COM1
26. Persephonella marina
27. Pyrolobus fumarii
28. Hyperthermus butylicus
29. Thermotoga petrophila
30. Thermotoga sp
31. Thermotoga neapolitana
32. Thermotoga naphthophila
33. Thermotoga maritima
34. Pyrolobus fumarii
35. Hyperthermus butylicus
36. Aeropyrum pernix
37. Thermotoga petrophila
38. Thermotoga sp
39. Thermotoga neapolitana
40. Thermotoga naphthophila
41. Thermotoga maritima

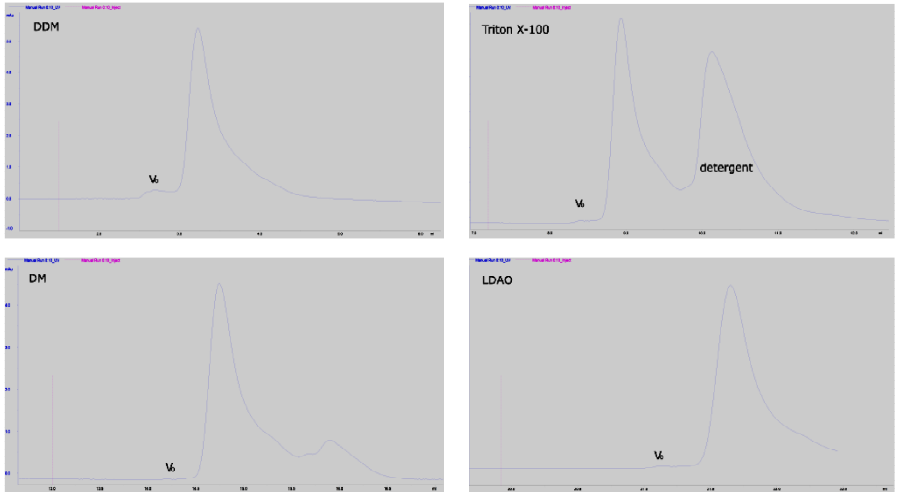

Identity
430
440
450
460
470
480
490
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

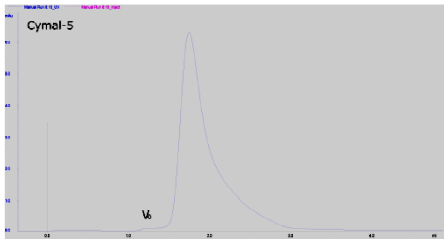
Appendix 2 Enzymatic assays of IPCT/DIPPS mutants. The IPCT/DIPPS activities were assessed in cell extracts of *E. coli* C43, harboring wild-type or mutant variants of the *ipctIdipps* gene from *A. fulgidus*. The cell pellets were suspended in 50 mM Bis/Tris/Propane-HCl (pH 8.0) containing 5% glycerol and 100 mM NaCl; cells were disrupted in a French press and debris removed by centrifugation (10,000*g*, 15 min, 4°C). Reaction mixtures (0.3 ml) containing 25 mM Bis/Tris/Propane-HCl (pH 8.0), 10 mM $MgCl_2$, 2.7 mM CTP, 7 mM *myo*-inositol-1-phosphate, and cell extract (0.9 mg of total protein) were incubated at 85°C for 30 min. The reactions were stopped on ice prior to the addition of deuterated water (200 µl) and 25 mM EDTA (pH 8.0). The reaction products of IPCT and DIPPS, *i.e.*, CDP-*myo*-inositol and DIPP, respectively, were quantified by ^{31}P-NMR using diglycerol-phosphate as an internal concentration standard.

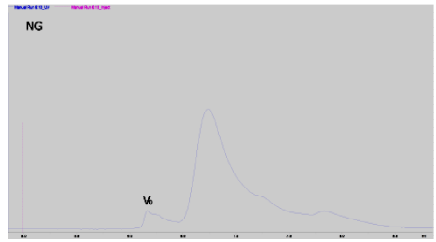
IPCT/DIPPS mutation	DIPP [mM]	%
WT IPCT/DIPPS	4.83	100
S308A	1.65	34.2
S308M	0.00	0
D357A	0.00	0
D357N	0.21	4.3

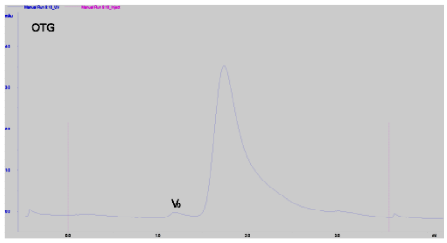
D360A	0.00	0
D360N	0.00	0
A364M	0.29	6.0
R365M	0.59	12.2
S371P	0.26	5.4
D378A	0.00	0
D378N	0.00	0
D382A	0.00	0
D382N	0.00	0
R443A	0.00	0
WT (no Mg^{2+})	0.00	0

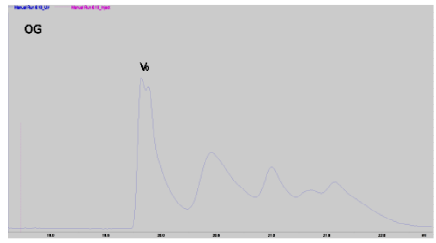
Appendix 3 Monodispersity tests of IPCT/DIPPS in the presence of different detergents. The samples were incubated for 6 days at 20˚C with tested detergent at about two times of its critical micelle concentration. Sample was analyzed by size exclusion chromatography (Superdex 200 5/150 GL). The detergents names abbreviations used: DDM - n-Dodecyl β-D-maltoside, DM - n-Decyl β-D-maltoside, LDAO - Lauryldimethylamine-N-Oxide, Cymal-5 - 5-Cyclohexyl-1-Pentyl-β-D-Maltoside, NG - n-Nonyl-β-D-Glucopyranoside, OTG - n-Octyl-β-D-Thioglucopyranoside, OG - n-Octyl-β-D-Glucopyranoside, Cymal-3 - 3-Cyclohexyl-1-Propyl-β-D-Maltoside, MEGA-8 - Octanoyl-N-methylglucamide, CHAPS - 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate, CHAPSO - 3-[(3-Cholamidopropyl)dimethylammonio]-2-Hydroxy-1-Propanesulfonate, C-HEGA-9 - Cyclohexylpropanoyl-N-Hydroxyethylglucamide, HEGA-8 - Octanoyl-N-Hydroxyethylglucamide, C-HEGA-8 - Cyclohexylethanoyl-N-Hydroxyethylglucamide, DTM - n-Decyl-β-D-Thiomaltopyranoside, Cymal-4 - 4-Cyclohexyl-1-Butyl-β-D-Maltoside, OTM - n-Octyl-β-D-Thiomaltoside, Fos-Choline-10 - n-Decylphosphocholine, Fos-Choline-9 - n-Nonylphosphocholine, Fos-Choline-8 - n-Octylphosphocholine, C12E8 - Dodecyl Octaethylene Glycol Ether, ZWITTERGENT 3-12 - n-Dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, DDAO - Decyldimethylamine N-oxide, ZWITTERGENT 3-10 - n-Decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, HEGA-10 - Decanoyl-N-Hydroxyethylglucamide, C8E5 - N-Octylpentaoxyethylene, NTG - n-Nonyl-β-D-Thioglucopyranoside, HTG - n-Heptyl-β-D-Thioglucopyranoside.

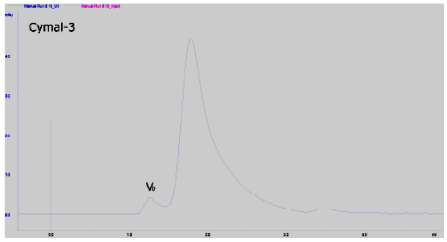


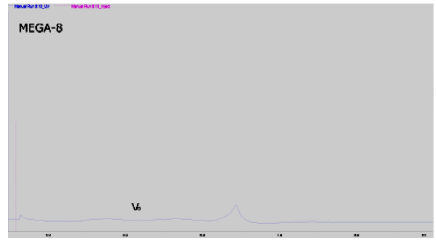

Cymal-5
V6

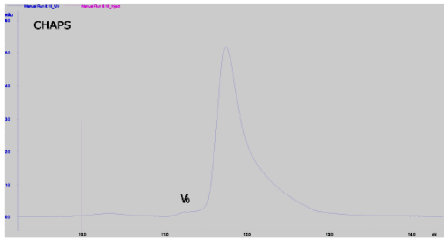

NG
V6

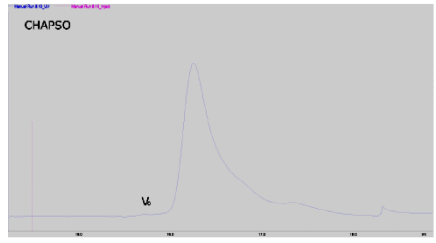

OTG
V6

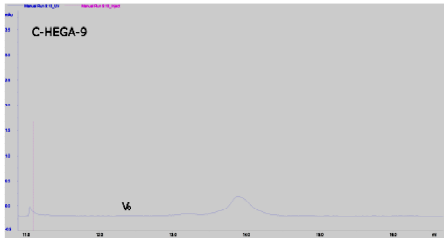

OG
V6

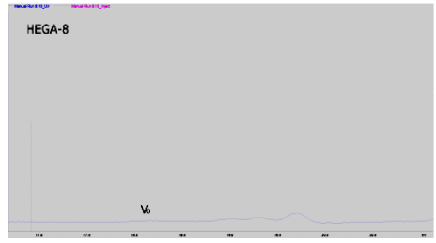

Cymal-3
V6

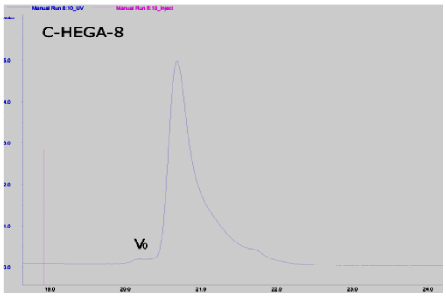

MEGA-8
V6

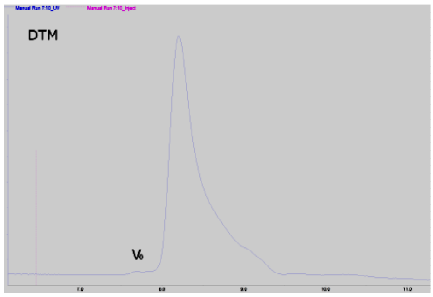

CHAPS
V6

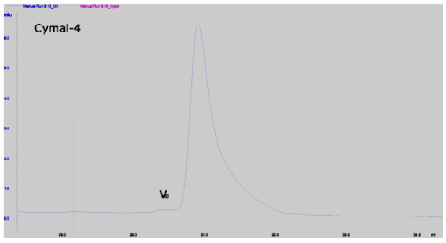

CHAPSO
V6

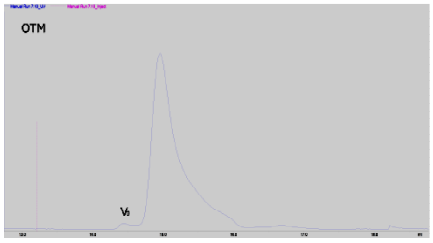

C-HEGA-9
V6

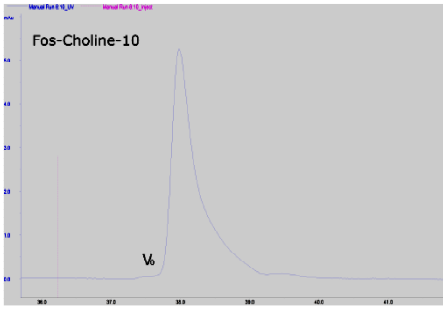

HEGA-8
V6

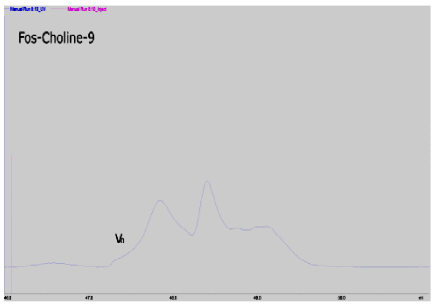

C-HEGA-8
V0

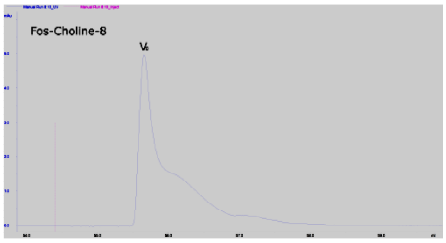

DTM
V0

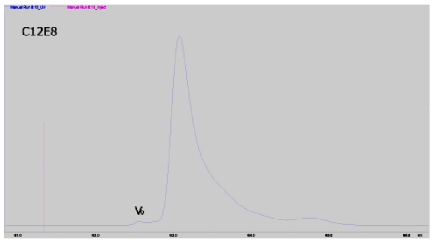

Cymal-4
V0

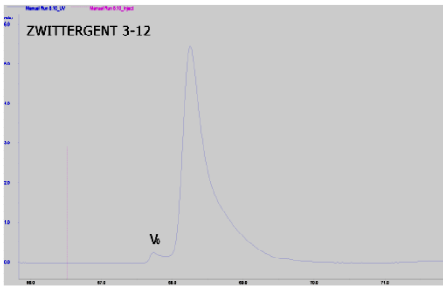

OTM
V0

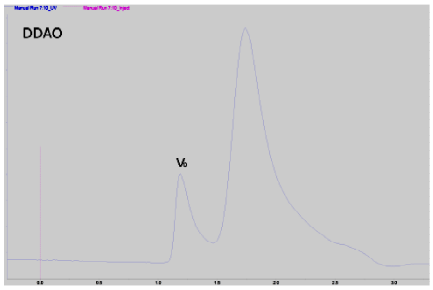

Fos-Choline-10
V0

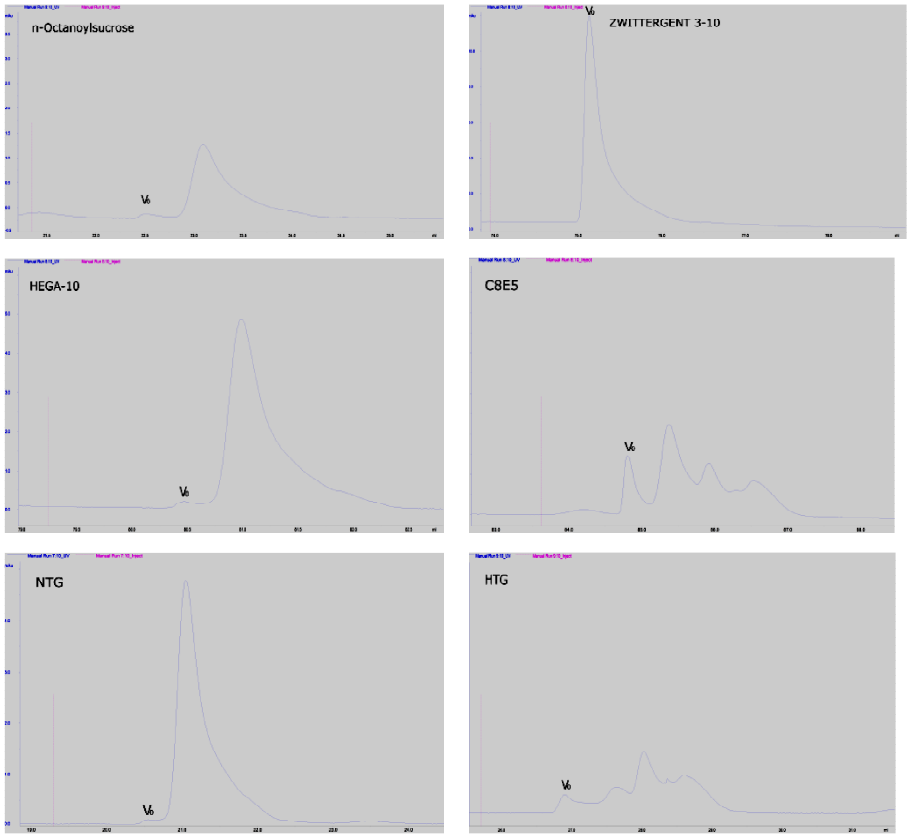

Fos-Choline-9
V0


Fos-Choline-8
V0


C12E8
V0


ZWITTERGENT 3-12
V0


DDAO
V0



REFERENCES

Gonçalves, L.G., Borges, N., Serra, F., Fernandes, P.L., Dopazo, H., Santos, H., 2012. Evolution of the biosynthesis of di-myo-inositol phosphate, a marker of adaptation to hot marine environments. Environ. Microbiol. 14, 691–701.

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

www.itqb.unl.pt