UNIVERSITAT DE BARCELONA

Insight into the structure and function of engineered biocatalysts: serine hydroxymethyltransferase from *Streptococcus thermophilus* and halohydrine dehalogenase D2 from *Gammaproteobacterium*

Giovanna Petrillo

UNIVERSITAT DE BARCELONA (UB)
Faculty of Pharmacy and Food Sciences

INSTITUTE OF ADVANCED CHEMISTRY OF CATALONIA(IQAC)
Department of Biological Chemistry and Molecular Modelling

MOLECULAR BIOLOGY INSTITUTE OF BARCELONA (IBMB)
Structural Biology Unit (SBU)

BIOCHEMIZE S.L.

Insight into the structure and function of engineered biocatalysts:

serine hydroxymethyltransferase from *Streptococcus thermophilus* and halohydrine dehalogenase D2 from *Gammaproteobacterium*

Giovanna Petrillo

2017

UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Industrial PhD Program in Biotechnology 2013-2014.
Departement of Biochemistry and Molecular Biology,
Faculty of Pharmaceutics, University of Barcelona (UB)

Insight into the structure and function of engineered biocatalysts: serine hydroxymethyltransferase from *Streptococcus thermophilus* and halohydrine dehalogenase D2 from *Gammaproteobacterium*

Doctoral thesis presented by Giovanna Petrillo, B.S degree in Biotechnology, for the degree of Industrial PhD from University of Barcelona.

PhD student:

Giovanna Petrillo

Thesis supervisor:
Prof. Pere Clapés Saborit

Thesis co-supervisor :
Prof. Isabel Usón Finkenzeller

Thesis tutor:
Prof. Josefa Badia Palacín

Thesis company tutor:
Dr. Jaume Mir Martinez

Barcelona, 2017

Ai miei genitori e ai miei fratelli,

colonne portanti

ad Alfo,

mattoni

di questa costruzione, che è la mia vita.

Acknowledgements

First, I would like to thank my supervisors Prof. Pere Clapés from the Institute of Advanced Chemistry of Catalonia (IQAC) and Prof. Isabel Usón, from the Institute of Molecular Biology of Barcelona (IBMB), without whom none of this would have been possible. I thank them for all their help and guidance regarding this whole process and for their patience and dedication.

I would also like to thank my company tutor, Dr. Jaume Mir Martinez, CEO of Biochemize.

I would like to thank all the members of my lab, Massimo and Claudia for their support and enriching discussion, and all the new and old members I met over these years. Not forgetting the people from CRI3, CRI6 and CRI7: their suggestions and encouragement were valuable contributions to my project. I would like to thank Anna Cuppi, my collegue, flatmate and friend, who has always been a great support. Great thanks go to Karel and Raquel from Biotransformation and Bioactive molecules group of IQAC. They gave me a lot of support with complicated chemical “stuff”.

Ringrazio Ivan, che mi ha regalato un’opportunità, tirandomi fuori dal grigiore e riservandomi un posto al sole di Barcellona e della sua amicizia.

Ringrazio, uno ad uno gli "Amigos de D10S", la mia famiglia qui a Barcellona.

Ringrazio te, Alfo, con il quale questa avventura è iniziata, con il quale gran parte delle cose belle della mia vita sono iniziate… dovunque fioriranno le viole, affonderano sempre le loro radici in un terreno coltivato con sincerità, rispetto e amore.

Desidero, infine, ringraziare la mia famiglia per l’appoggio incondizionato e la stima, che sono un flusso incessante, come il mare.

Contents

Abstract

The rapid progress in structural and molecular biology over the past fifteen years has allowed chemists to access the structures of enzymes, of their complexes and of engineered variants. This wealth of structural information has led to a surge in the interest in enzymes as elegant chemical catalysts of industrial application.

The study of structure-function of enzymes is central for designing active-site-directed mutations that expand the catalytic repertoire of enzymes properties.

In this study, the structure of two enzymes interesting for different biocatalytic applications was determined in order to shed light on their catalytic reaction mechanism and guide protein engineering experiments.

The crystal structure of a novel serine hydroxymethyltransferase from *Streptococcus thermophilus* (SHMT) at 2.2 Å resolution, in the unliganded form and in complex with glycine and L-threonine was determined. Furthermore, two active-site variants have been functionally characterized in complex with natural and non-natural substrates. No active site catalytic residues were revealed, and a structural water molecule was assumed to act as catalytic base in the retro-aldol cleavage reaction. Moreover the X-ray analysis of the variants complexes explains the molecular bases of their enhanced activity and broad substrate tolerance.

Abstract

Halohydrin dehalogenase D2 from Gammaproteobacterium was also studied and its structure was solved at 1.6 Å resolution using a new method, Arcimboldo_Shredder, that exploits fragments from distant homologs to phase macromolecular structures.

A similar structure and catalytic mechanism to previously reported protein of the same family was revealed. These results broaden the insight into the structural determinations that govern the reactivity and selectivity in the haloalcohol dehalogenase family and open the possibility to a rational re-design of them.

Key words: crystallization, X-ray diffraction, structural biology, aldolases, serine hydroxymethyltransferase, biocatalysis, halohydrin dehalogenases, enzymes.

Frequently used abbreviations

Å	Angstrom
A.U.	Asimmetric Unit
Da	Dalton
DLS	Dynamic Light Scattering
3D	Three-dimensional
HCHO	Formaldehyde
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2- ethylenesulfonic acid]
HHDH	Halohydrin dehalogenases
HPLC	High performance liquid chromatography
H4PteGlu	Tetrahydropteroylglutamate
IPTG	Isopropyl-1-thio- -D-galactopyranoside
LTA	L-threonine aldolase activity
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PLP	Pyridoxal 5'-phosphate
K_m	Michaelis-Menten constant
Rmsd	Root mean square deviation
SDS	Sodium dodecylsulfate
SHMT	Serine hydroxymethyltransferase
$SHMT_{Sth}$	SHMT from *Streptococcus thermophilus*

***bs*SHMT** SHMT from *Bacillus stearothermophilus*

***e*SHMT** SHMT from *Escherichia coli*

THF Tetrahydrofolate

WT *wild-type*

$SHMT_{Sth}$ structure identification codes

WT*apo* structure of $SHMT_{Sth}$ *wild-type apo*-form
WTgly structure of $SHMT_{Sth}$ *wild-type* glycine external aldimine
WTLthr structure of $SHMT_{Sth}$ *wild-type* L-threonine external aldimine
Y55SLThr structure of $SHMT_{Sth}$ Y55S L-threonine external aldimine
Y55SDThr structure of $SHMT_{Sth}$ Y55S in complex with D-threonine
Y55SPLS structure of $SHMT_{Sth}$ Y55S D-serine external aldimine
Y55SDSerPLP structure of $SHMT_{Sth}$ Y55S internal aldimine with D-serine in the active site
Y55TPLS/DSerPLP structure of $SHMT_{Sth}$ Y55T L-threonine external aldimine

Proteinogenic amino acids three- and one- letter codes

Glycine **Gly G**

Alanine **Ala A**

Valine **Val V**

Leucine **Leu L**

Isoleucine **Ile I**

Phenylalanine **Phe F**

Proline **Pro P**

Methionine **Met M**

Tryptophane **Trp W**

Histidine **His H**

Threonine **Thr T**

Cysteine **Cys C**

Tyrosine **Tyr Y**

Glutamine **Gln Q**

Asparagine **Asn N**

Aspartate **Asn D**

Glutamate **Glu E**

Arginine **Arg R**

Lysine **Lys K**

General Introduction

1. Enzymes, a brief overview

1.1. Enzymes and their early use in biotechnology

In living organisms most of the reactions are catalysed by protein molecules called “enzymes”, which are the catalytic machinery of living systems. They catalyse the vast repertoire of chemical reactions found in nature and are probably the most studied of biological molecules. They accelerate biochemical reactions, lowering the activation energy compared to the corresponding uncatalysed reactions thus resulting in higher reaction rate at the same temperature. In this way catalysts permit reactions or processes to take place more effectively or under milder conditions than would otherwise be possible.

The first enzyme to be discovered was amylase in 1833, which was originally isolated from barley and catalyses the conversion of starch into sugars. This discovery triggered the development of commercial bread-making and brewing techniques as well as the development of fermentative enzymes at the beginning of the 20th century. Then urease, an enzyme that catalyses the hydrolysis of urea into ammonia and carbamate, was isolated and crystallised in 1926 (Sumner, 1926). These findings led to the demonstration that enzymes are actually proteins, a topic of controversy at that time. After forty years the first three-dimensional structure of an enzyme was determined using X-ray diffraction techniques. In 1965, Phillips and co-workers obtained the structure of the hen egg-white lysozyme, which hydrolyses peptidoglycan in bacterial cell walls (Blake et al., 1965).

The ability of enzymes to perform biochemical catalysis has two main characteristics: acceleration of reaction rate and specificity (Fersht, 1999, Silverman, 2002). In fact enzymes are able to speed up chemical

reactions by lowering their activation energy. In general, the lower the activation energy of a reaction has, the faster the rate of reaction is. Moreover, enzymes are often highly selective for their substrates and the reactions they catalyse.

1.2. Enzyme chemistry: catalytic site, mechanism and cofactors

Although enzymes are often large molecules comprising many hundreds of amino acids, the functional regions of an enzyme are generally restricted to clefts on the surface that comprise only a small part of the enzyme's overall volume (Gutteridge and Thornton, 2005). The structural studies revealed that the active site, where enzymatic catalysis takes place, are cracks or hollows on the surface of the enzyme caused by the way the protein folds itself up into its tertiary structure. Molecules of the precise right shape, electronic structure and with precisely the right arrangement of attractive groups can fit into this active site. As known for a long time, the active site is constantly being reshaped by its interaction with the substrate, as suggested by Koshland in 1958 (Koshland et al., 1958). Substrate does not bind to an active site as if it were specifically the right shape, but the amino acid side-chains, which are a part of the active site, are modelled into a specific position. This position allows the enzyme to start the catalysis process. The Koshland's model has been called the "induced fit" theory (**Fig.1**).

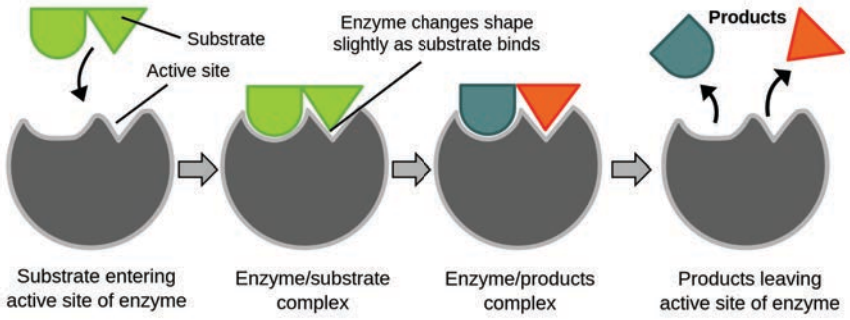


Figure 1: "Induced fit" model

When the substrates enter in the active site, the enzyme changes its shape slightly to fit the substrates further. Once the product is produced and released, the enzyme goes back to its original shape.

Experimental evidence from X-ray crystallography suggested a model to explain how this process takes place: the enzyme and substrates initially form a complex in the active site, which may induce large conformational changes in their structures.

First suggested by Linus Pauling, substrate binding is followed by a stabilisation of a highly ordered transition state. Enzymes decrease the activation energy of the reaction because they are complementary in shape and electrostatic properties to the rate-limiting transition state, which explains the vast rate acceleration compared to un-catalysed reactions (Pauling, 1946). One or more reaction intermediates are generated, which then develop into products and are finally released from the active site. As the substrate approaches the enzyme, molecular recognition takes place and the complex enzyme-substrate arises from interactions of the substrate with various amino acids in the active site. There are two types of interactions driving the complex formation: covalent interactions, which involve the sharing of electrons

within a strong chemical bond; and non-covalent interactions comprising electrostatic, dipole, hydrogen bonding, hydrophobic and van der Waals forces. Non-covalent interactions also contribute to enzyme catalysis by stabilising the transition state and destabilising the ground state. Although individually weak, they collectively make a strong interaction that is maximal when the transition state is formed. In addition, their reversibility also allows the product to be released from the enzyme active site (Silverman, 2002). Despite the vast research aimed at building basic principles of enzyme catalysis, we still know very little about certain aspects of this phenomenon.

The active site of an enzyme comprises about ten to twelve amino acids, the "catalytic toolkit". About three or four of them, known as catalytic amino acids, are directly involved in the catalysis of the biochemical reaction and form the catalytic site (Gutteridge and Thornton, 2005). Their role is defined according to the specific chemical function they performed in the mechanism. In general, catalytic amino acids tend to be conserved, they show slight preference to be located in coiled-coil regions of secondary structure and they exhibit limited solvent accessibility (Bartlett et al., 2002). Histidine, cysteine, lysine, serine and aspartate are the amino acids often engaged in catalytic activities, whereas aliphatic amino acids such as alanine, leucine and glycine are rarely involved (Holliday G. L. et al., 2007). Their most common functions are transition state stabilisation, general acid/base (proton donor and acceptor) and nucleophilic covalent catalysis (Bartlett et al., 2002, Holliday G. L. et al., 2007). In multiple occasions residues act in combination forming catalytic units, for instance some residues may polarise the substrates or orientate other residues to maximise the probability that catalysis occurs (Gutteridge

and Thornton, 2005). Enzymes use several types of mechanisms to convert substrates into products. More than two-thirds of enzyme reactions rely on acid/base chemistry, especially proton transfer reactions (Holliday G. L. et al., 2007). The second most common mechanism is nucleophilic catalysis such as substitutions, additions and eliminations, whereas electrophilic reactions are rare. To gain further insight into the enzyme mechanism two key parameters must be considered: the Michaelis-Menten constant (K_M) and the catalytic rate constant (k_{cat}) (Stitt and Gibon, 2014, Wittig et al., 2014). In general, enzyme reactions are described by Michaelis-Menten kinetics where K_M captures the substrate binding affinity to the enzyme and links the reaction rate with the substrate concentration. Secondly, k_{cat} captures the speed whereby the substrate turns into product in active site of the enzyme (Cornish-Bowden, 2014). Studies comparing k_{cat} with the rate constant of the uncatalysed reaction (k_{uncat}) strongly support the idea of acceleration of reaction rate achieved in enzyme reactions. Some enzymologists prefer combining these two parameters in a ratio known as the specificity constant (k_{cat}/K_M), which measures the enzyme's ability to discriminate among competing substrates. Nevertheless when k_{uncat} data are available, the so-called catalytic proficiency ($(k_{cat}/K_M)/k_{uncat}$) is also common in the literature (O'Brien and Herschlag, 1999).

The enzymatic catalysis of many biochemical reactions requires the presence of cofactors. A cofactor is an organic molecule or metal ion that binds to the active site and is essential for catalysis. The use of cofactors allows enzymes to expand the catalytic abilities achieved by the 20 naturally occurring amino acids (Holliday G. L. et al., 2007), yet not all enzymes require a cofactor. Organic cofactors are mainly

nucleotides, amino acids or fatty acids substructures. Although they chemically resemble other metabolites in the cell, cofactors are on average significantly more polar and slightly larger (Fischer et al., 2010).

1.3. Enzyme selectivity

Enzymes are exquisitely selective and chemically effective catalysts. These two properties, selectivity and chemically efficiency, distinguish enzymes from ordinary chemical catalysts (Hedstrom, 2001).

Enzymes display four major types of selectivity:

- Substrate selectivity (chemoselectivity), which is the ability to distinguish and act on a subset of compounds within a larger group of chemically related compounds;
- Stereoselectivity, which is the ability to act on a single enantiomer or diastereomer exclusively;
- Regioselectivity, which is the ability to act exclusively on one location in a molecule;
- Functional group selectivity, which is the ability to act on one functional group selectively in the presence of other equally reactive or more reactive functional groups.

The selectivity of enzymes makes them powerful tools for asymmetric synthesis, fuelled by the increasing demand for optically pure synthetic intermediates in pharma industries (Bornscheuer et al., 2012).

2. Enzymes and its industrial applications

2.1. Biocatalysis for industrial production of chemicals

The chemical industry has a central role in modern society; since it provides society with a large number of value-added products and is one of the biggest economic sectors worldwide (Schmid and Smith, 2002). More recently, the conventional chemical industry has been forced to innovate in order to successfully penetrate already saturated markets (Ferrer et al., 2012). This has resulted in increasing focus on production of chemicals from renewable sources, promoting greener synthetic routes and generating less toxic by-products and waste, without compromising the product quality. Green Chemistry, defined as "the design of chemical products and processes to reduce or eliminate the use and generation of hazardous substances"(Anastas and Eghbali, 2010), has been promoting the design of next generation processes and products, by providing guidelines for environmentally friendly and economically competitive processes. Using alternative substrates, energy sources and innovative synthetic routes, bioprocesses have brought many innovations to the polymer, biofuel, textile, food, health care and pharmaceutical industries, amongst others (Frazzetto, 2003). Bioprocesses have provided many innovative routes for the chemical industry.

Bioprocesses can be classified into fermentation and biocatalysis. Fermentation refers to the use of growing cells to make the product of interest. Biocatalysis may broadly be defined as the use of enzymes, which can be crude extracts, purified enzymes, or whole-cells. Biocatalytic processes deal with the organic synthesis of chemicals (small molecules) such as building blocks for added-value chemicals, amino acids, agrochemicals and active pharmaceutical ingredients

(APIs). In the past decades, enzyme and whole-cell biocatalysis have been applied in the production of various chemicals and proof of such significant growth is the fact that the number of commercially available enzymes isolated from different biological sources has increased significantly (Rozzell, 1999). The exquisite selectivity of enzyme catalysed reactions, yielding single stereoisomers with few side reactions and easier separation of products (Meyer *et al*., 1997) under mild reaction conditions (pH, temperature, often aqueous conditions, etc.), make biocatalytic processes well positioned to contribute to greener industrial processes, in line with the 12 Principles of Green Chemistry (Anastas and Eghbali, 2010). Enzymes found in nature operate at the host conditions, which are typically far milder than the conditions required for an industrial chemical processes, such as high substrate and product concentrations, "unnatural" substrates, presence of reaction additives (e.g. solvents), among others.

Hence, often biocatalytic processes do not meet the required process metrics that are key for an economically feasible process at an industrial scale (high concentration, high yield, high space-time yield and high biocatalyst yield) (Tufvesson P. et al., 2011).

3. Biocatalyst engineering

Nature has supplied us with a vast array of biocatalysts capable of catalysing numerous biological reactions. Unfortunately, naturally occurring biocatalysts are often not optimal for many specific industrial applications. Moreover, naturally occurring biocatalysts may

not catalyse the reaction with the desired non-natural substrates or produce the desired products (Johannes et al., 2005).

To address these limitations, molecular techniques have been developed to create improved or novel biocatalysts with altered industrial operating parameters.

Therefore, in order to improve the catalytic performance of the industrial enzymes protein, engineering plays a major role, resulting in the reduction of cost and time of processing in the respective industries. Changing cofactor specificity, discovering paths of molecular evolution, designing more powerful and efficient enzymes, locating active sites and crucial residues, shifting the substrate specificity toward new substrate, improving binding affinity, changing in pH tolerance and inhibitor tolerance, improving solubility and resistance to extreme conditions, are some of the modifications that can be achieved by *in vitro* protein engineering. In particular, enzymes can be successfully altered by modifying the active site to optimize their efficiency and selectivity. Altering the ligand binding sites, incorporation of new functional groups into the active site, and site-directed mutagenesis are some of the approaches that modify the active site of an enzyme.

4. Structural enzymology

Understanding the molecular architecture and catalytic mechanism underlying enzyme function in living cells from bacteria to humans is of fundamental importance in order to rationally modify the enzyme to modulate its catalytic properties.

In attempting to engineer novel biocatalysts, the catalytic centers of existing enzymes provide obvious starting points. Even small changes

in the amino acid composition can account for their different activities, which supports the hypothesis that a wide variety of enzymatic activities divergently evolved from a limited set of primordial enzymes (Yoshikuni et al., 2006). Such activity switching demonstrates that the inherent structural plasticity and catalytic versatility of enzyme active sites can give rise to new biocatalysts. Thus, productive searching of interesting mutations in the enzyme active sites is of a great interest to enhance "promiscuous" activities and thereby increase their effectiveness in synthetic applications. Mutation of active-site residues often dramatically changes the properties of an enzyme. Most changes of this type are simply inactivating (Peracchi, 2001); sometimes, however, the altered enzymes have broadened substrate specificities or altered reaction profiles from which novel activities emerge (Penning and Jez, 2001). The prediction and control of these effects provides a promising way to access new functions.

How then can one efficiently identify interesting changes? The use of sequence, structural, and mechanistic information has proven to be a productive basis for designing active-site-directed mutations that expand the catalytic repertoire of enzymes. This strategy relies on targeted substitutions that do not disrupt the global protein fold, allowing (rational) exploitation of the unchanged residues in the context of the "new" active site (Toscano et al., 2007).

To study the function and molecular structure of enzymes, we employ biochemistry and X-ray crystallography, along with other structural methods, from single-particle electron microscopy to small-angle X-ray scattering.

Through the use of these techniques, researchers are able to probe mechanism and specificity to gain a greater understanding of how

enzymes work and how they function in the context of molecular pathways in the cell. Such knowledge could provide the basis for new medical treatments, pollution-control strategies, and fuel production, among many other applications.

4.1. X-ray crystallography as a method to design new biocatalysts

X-ray crystallography is currently the most widespread technique for structure determination in atomic detail of proteins and biological macromolecules. Increasingly, those interested in all branches of the biological sciences require structural information to shed light on previously unanswered questions. Furthermore, the availability of a protein structure can provide a more detailed focus for future research. X-ray crystallography is one of the most detailed "microscopes" available today for examining macromolecular structures. The three-dimensional structures of biological macromolecules are often considered as "gold standard" of data describing the molecular architecture of important proteins. When the structure of interest have been determined, macromolecular models can yield a wealth of information necessary for rational drug design, site directed mutagenesis, elucidation of enzyme mechanisms, and specificity of protein–ligand interactions are just a few of the areas in which X-ray crystallography has provided clarification.

During the past two decades, we have witnessed an unprecedented success in the development of highly potent and selective drugs or lead compounds based on information obtained from the crystal structures of target proteins. Prominent examples include transition-state analog inhibitors for influenza virus neuraminidase (Feng et al., 2012) and,

perhaps most notably, inhibitors of HIV protease (Srivastava et al., 2012). Moreover, crystal structures inspire new hypotheses and experiments to probe biological macromolecules regarding molecular mechanisms, plausible binding modes, and the feasibility of small molecule agents to serve as scaffolds for lead compounds

Objectives

The following thesis can be divided in two related parts:

CHAPTER 1 Structure and catalytic mechanism elucidation of serine hydroxymethyl transferase from *Streptococcus thermophilus* ($SHMT_{Sth}$) and variants

.

CHAPTER 2: Structural solution of halohydrin dehalogenases D2 (HheD2) from *Gammaproteobacterium.*

The enzymes presented in this work are part of a 48-month collaborative project funded by the European Commission under the Horizon 2020 Programme, called Carbazymes.
During its course, the project will implement the biocatalytic synthesis of Active Pharmaceutical Ingredients (APIs) and bulk chemicals by using a broad platform of unique C-C bond-forming enzymes, opening the doors to the development of biocatalytic routes for the synthesis of these high demand molecules and pioneer environmentally friendly processes for much needed products, which are particularly relevant for Europe´s sustainable future and competitive ability in the global industry. The structural studies of promising enzymes in order to elucidate their catalytic mechanism and offer a rational basis to design improved or customized catalytic properties constitute our role in the frame of this project.

Each of this chapter contemplates differentiated objectives:

CHAPTER 1

I. The main goal of the project discussed in this chapter is the crystallisation and structural determination of serine hydroxymethyltransferase from *Streptococcus thermophilus* ($SHMT_{Sth}$) by X-ray crystallography. This enzyme is a pyridoxal-5'-phosphate, found to catalyse aldol addition of glycine to aldehydes with good stereoselection under kinetic control.

III. A second goal is to crystallise the most interesting variants from a synthetic point of view, in complex with different reaction substrates and thus characterise the three-dimensional structure of reaction intermediates.

II. Finally, the elucidation of the catalytic mechanism at atomic resolution and all the species involved.

CHAPTER 2

I. The main goal of the project discussed in this chapter is the crystallisation of the halohydrin dehalogenase D2 from *Gammaproteobacterium* (HheD2).

II. The second goal is to determine the three-dimensional structure of HheD2 from *Gammaproteobacterium* by X-ray crystallography, using a new method, Arcimboldo Shredder that exploits fragments from distant homologs to phase macromolecular structures. Molecular replacement as a method to use previous structural information to

solve the phase problem requires a low root mean square deviation (r.m.s.d.), which usually correlates with the existence of homologs with 30% or more sequence identity. In HheD2, the goal was to extend the radius of convergence of conventional molecular replacement, solving the structure from the available homologs of lower identity.

II. Finally, the elucidation of overall structure and the catalytic site with all the species involved at atomic resolution.

CHAPTER I

Structure and catalytic mechanism elucidation of serine hydroxymethyl transferase from *Streptococcus thermophilus* and variants

Introduction

1.1. Pyridoxal 5'-phosphate (PLP) and PLP-dependent enzymes

Pyridoxal 5'-phosphate (PLP) has been shown to act as a coenzyme in a large and varied range of biochemical reactions. PLP is part of the vitamin B6 group, which consists of 6 vitamers: pirydoxal, pyridoxine and pyridoxamine and their corresponding phosphorylated forms (**Fig.2**) and it is used by a variety of enzymes in all organisms (Percudani and Peracchi, 2003). Some organisms, like bacteria, fungi and plants, can carry out *de novo* synthesis of vitamin B6, but not humans, that have to acquire it exogenously from their diet and enzymatic salvage pathways.

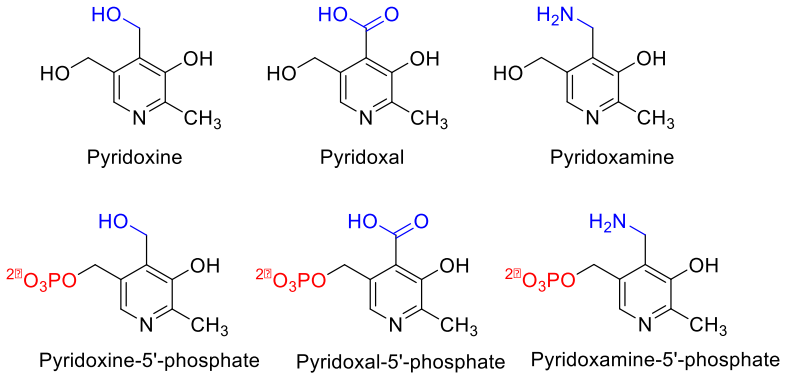


Figure 2: Structure of B6 vitamers

PLP structures and related vitamers. From left to right: pyridoxal 5'-phosphate, the active form of vitamin B6 used as enzyme cofactor; pyridoxal (PN), the vitamin B6 form that is most commonly given as dietary supplement; pyridoxamine 5'-phosphate (PMP), the other natural occurring vitamin B6- derived catalyst; and their corresponding phosphorylated forms.

Following its identification in 1951 as one of the active vitamers of vitamin B6 (Heyl et al., 1951), PLP has been the subject of extensive research directed toward understanding its catalytic versatility. Nowadays, the basic mechanism of PLP-assisted reactions has been well characterized in detail.

There are more than 160 chemical reactions catalysed by PLP, which can be divided according to the position at which the net reaction occurs. Reactions at the α-position include transamination, decarboxylation, racemization, elimination and replacement by an electrophilic group. Those at the β- or γ- position include elimination or replacement (**Fig.3**)

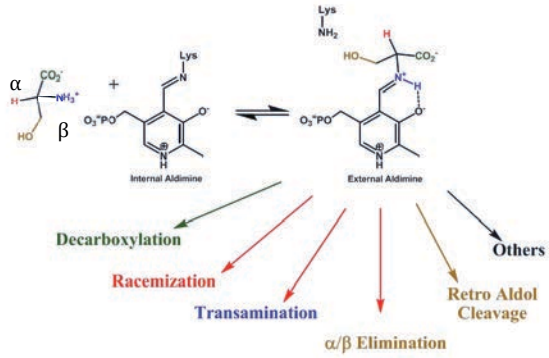


Figure 3: Variety of reactions catalysed by PLP

The most important reactions catalysed by PLP-dependent enzymes are shown here. Several other reactions are omitted for clarity. The amino acid, taken as example in the scheme, is serine and the carbons α and β group are marked. The picture is reproduced from (Toney, 2005).

In addition to their catalytic versatility, PLP-dependent enzymes have attracted attention because of their widespread involvement in cellular processes. Almost all PLP-dependent enzymes are associated with biochemical pathways that involve amino compounds, predominantly amino acids and amino acid-derived metabolites, but they are also found in the biosynthetic pathways of amino sugars (He and Liu, 2002) and other amine-containing compounds. Moreover they have been identified as drug targets; for example SHMT has been identified as target for cancer therapy (Snell, 1984), as further explained later in this chapter.

The interest for these enzymes was supported by their availability in large amounts of protein available from organs like liver and heart, the existence of mitochondrial and cytosolic isoforms, the availability of natural mutants, the complex chemistry of the catalytic mechanism, and the spectral changes of the coenzyme associated to the enzyme catalysis. Today, PLP-dependent enzymes are still a subject of many investigations, further triggered by the outcome of functional genomics, the improvement of cloning, expression and purification techniques, and the discovery of their involvement in several human diseases, thus calling for specific inhibitors/drugs.

1.2. Structural diversity within PLP-dependent enzymes with emphasis on fold type I

PLP-dependent enzymes are a large ensemble of biocatalysts that make use of the same cofactor but have distinct evolutionary origins and protein architectures (Eliot and Kirsch, 2004).

According to their three-dimensional structure, PLP-dependent enzymes are grouped into five evolutionarily unrelated superfamilies, corresponding to as many different fold types, type I, II, III, IV and V (Grishin et al., 1995). These families have been named from their more representative enzyme. Fold-type I represents the largest, functionally most diverse and best characterized group. It is found in a variety of aminotransferases and decarboxylases, as well as in enzymes that catalyse α, β or γ eliminations. Fold-type II is found mainly in enzymes that catalyse α, β or γ eliminations reactions. Fold-type III is found in alanine racemases and in a subset of amino acid decarboxylases. Fold type IV enzymes include D-alanine aminotransferase and a few other enzymes. Finally, the fold-type V group includes glycogen and starch phosphorylases.

Enzymes in the fold-type I, aspartate aminotransferase family, invariably function as homodimers in prokaryotes or higher-order oligomers in eukaryotes. The active site is located at the interface of the domains and is delimited by amino acid residues contributed by both subunits of the dimer. There are two active sites per dimer and, in general, they work independently. Each monomer folds into two domains: a large N-terminal domain and a small C-terminal domain. The central feature of the large N-terminal domain is a seven-stranded β-sheet. The small domain at the C-terminal part of the chain folds as a three- or four-stranded β-sheet partly covered with helices on one side.

PLP is covalently attached to the large domain via the ε- amino group of a lysine residue at the N terminus of a short helix following a β strand. The phosphate is anchored to the N terminus of another α helix on the other side of the β sheet, so that the aromatic ring of PLP packs against the neighbouring β strands. The active site is located in a cleft between the two domains, at the interface between the two subunits of the dimer. Residues from both domains and both subunits are involved in cofactor binding.

The structures of fold-type II enzymes, referred as tryptophan synthase family, are similar to those of fold-type I, but the proteins are evolutionary distinct (Metha et al., 1993). In contrast to type I, in the case of type II only residues from one monomer contribute to the composition of the active site (Hyde et al., 1988). Nevertheless, the functional form remains a homodimer or higher-order oligomer.

Fold type IV enzymes, D-amino acid aminotransferase family, are also functional homodimers and the catalytic portion of each monomer is participated by the small and large domains, where the cofactor is bound in the site that is specular with respect to the binding site in the fold types I and II. The *re* face (rather than *si* face) is exposed to the solvent (Gallagher et al., 1998).

Fold type III members, constituting the alanine racemase family are quite different from the other PLP-dependent enzymes. They consist of a classical α-β barrel and a second β-strain domain. Locally, the mode of binding PLP remains similar to that of other fold types, with the phosphate group anchored at the N-terminus of an α-helix, H-bond interactions made to the 3'-OH and the presence of a lysine Schiff base (Kern et al., 1999). Finally, fold-type V, glycogen phosphorylase family, enzymes show a difference in utilizing the phosphate group of

the cofactor for catalysis. Nevertheless, these enzymes are also obligate dimers, as residues from both monomers contribute to the active site. In **Fig.4** the three-dimensional structure of a representative enzyme for each fold-type is shown.

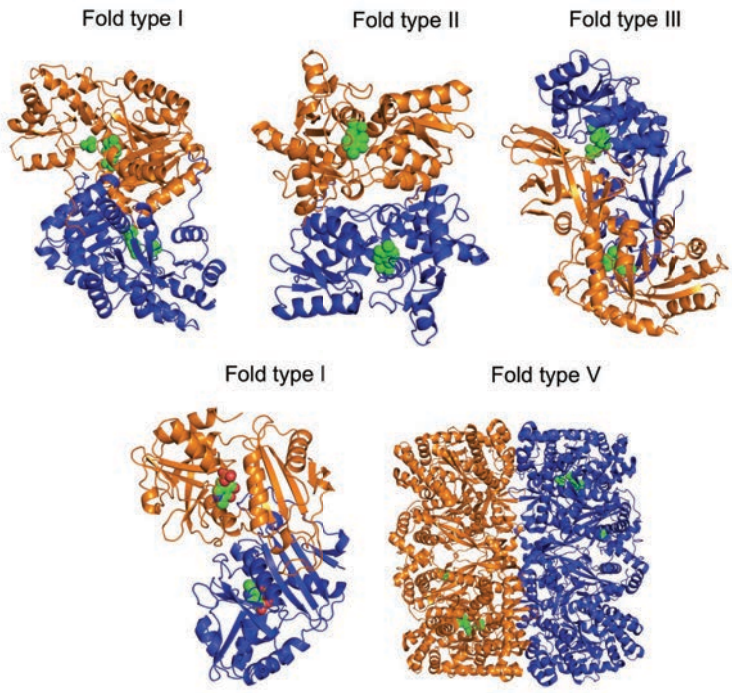


Figure 4: Ribbon diagram of representative enzymes for fold types I-V

Fold type I, *E.coli* aspartate aminotransferase is shown (PDB code: 1ASN). Fold type II, *Salmonella typhimurium* O-acetylse Fold type II, *Salmonella typhimurium* O-acetylserine sulfhydrylase (PDB code: 1OAS). Fold type III, Bacillus stearothermophilus alanine racemase (PDB code: 1SFT). Fold type IV, *Thermophilic Bacillus sp*. D-amino acid aminotransferase (PDB code: 1DAA). Fold type V, *Rabbit* glycogen phosphorylase (PDB code: 1GPA). In all the structures a homodimer is represented, with the individual monomers distinguished through a different colour. The cofactors are shown in green, marking the position of the active site in each structure.

Fold type I represents the largest sub-class of PLP dependent enzymes and also the most diverse class (Christen and Mehta, 2001). It includes the serine hydroxymethyl transferase (SHMT) family, which constitutes a small but still growing superfamily of enzymes, many of which play important biological roles as later summarised. This chapter will concentrate on the structural study of serine hydroxymethyltransferase from *Streptococcus thermophilus* in its atomic structure and mechanistic implications and the structure analysis of two interesting variants.

1.2.1 A common catalytic mechanism

Despite the wide range of chemical reactions catalysed by the PLP cofactor, a general catalytic mechanism remains consistent throughout the different fold-types.

Two interdependent chemical properties of PLP are determinant in catalysis. First, its ability to form imines with primary amino groups through its aldehyde group, and second its ability to stabilize negative charge development at Cα in the transition state that is formed after condensation of the amino acid substrate with PLP to form a Schiff base intermediate, as shown below in more detail.

Typically, this cofactor forms two types of imine: internal and external. In the free enzyme an internal aldimine is formed with an active lysine side chain. In fact, the first and common step for all PLP-dependent enzyme catalysed reactions is a Schiff base exchange reaction. Then the incoming amine-containing substrate displaces the lysine ε-amino group from the internal aldimine to form a new aldimine with the substrate, called external aldimine. External aldimines are formed with amino acids and closely related compounds

and constitute the common central intermediate for all PLP catalysed reactions. Divergence in reaction specificity occurs from this point on. (**Fig. 5**).

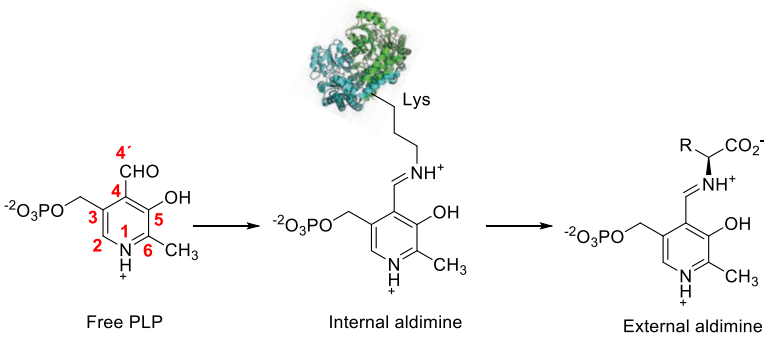
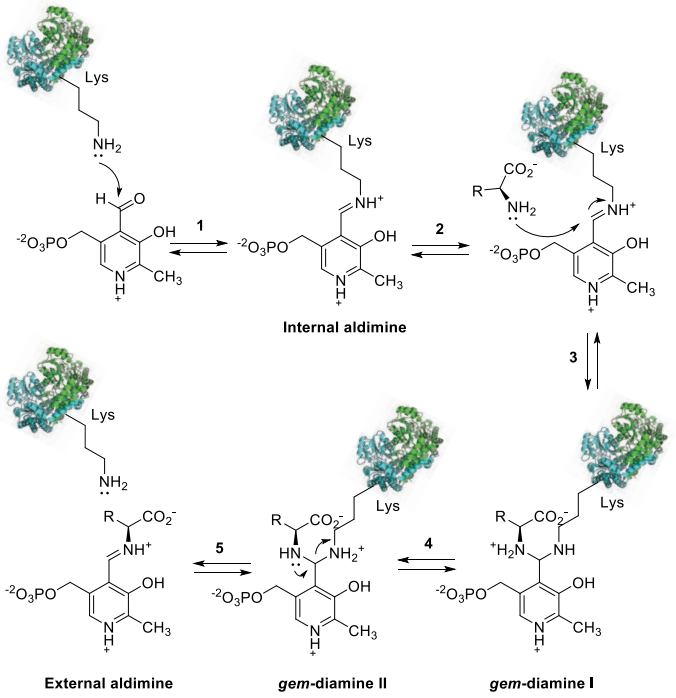


Figure 5: PLP imines

The PLP cofactor is able to form imines, binding the enzyme (internal aldimine) and the amino acid substrate (external aldimine).

Both types of imines react reversibly with primary amines (di Salvo et al., 2013a) in a transaldimination reaction, with formation of a geminal diamine intermediate, to allow binding of substrates and release of products (**Scheme 1**).



Scheme 1: General catalytic mechanism for PLP-dependent enzymes

PLP reacts reversibly with primary amines to form imines; (1) formation of the internal aldimine with the active site lysine residue; (2 to 5) transaldimination reaction: formation of the external aldimine with the substrate amino acid, through the formation of geminal diamine intermediates I and II. The transaldimination reaction between the internal and external aldimines allows substrate binding and product release.

In all cases, the cofactor stabilizes a negative charge at the Cα in the amino acid substrate at the transition state that is formed after addition of the amino acid to PLP furnishing the Schiff base.

1.3 Serine hydroxymethyltransferase (SHMT), a fold type I PLP-dependent enzyme

Serine hydroxymethyltransferase (SHMT) is a PLP-dependent enzyme that catalyses simultaneously the conversion of L-serine to glycine and tetrahydrofolate (THF) to N^5,N^{10}-MethyleneTHF (Schirch, 1982)

Tetrahydrofolates are present in cells as a family of metabolic cofactor that carry and chemically activate single carbons for a network of biosynthetic pathways referred to as folate-mediated one-carbon metabolism (Stover, 2004).

The currently accepted mechanism for the hydroxymethyltransferase reaction consists of a folate-dependent aldol cleavage via direct nucleophilic attack of *N*5 of H4PteGlu to Cβ of serine, which results in the formation of glycine and N^5,N^{10}-MethyleneTHF.

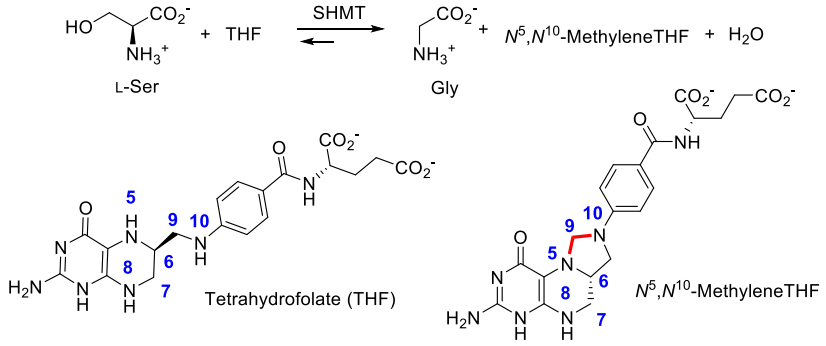


Figure 6: SHMT reaction

SHMT catalyses the inter-conversion of L-serine to glycine in the presence of THF. In the THF structure the numbers for the carbon atoms involved in the reaction are highlighted in blue.

SHMT belongs to the α-family of PLP enzymes a group of PLP enzymes that cleave one of the bonds at the α-carbon in their amino acid substrate. SHMT actually shares only very low sequence identity with other members of this family. To date, the amino acid substrates and sequences of SHMTs from 32 species have been deposited in the SWISSPROT data bank and exhibit sequence identities varying between 32% and 93% (Renwick et al., 1998) (Scarsdale et al., 1999).

Furthermore, SHMT is one of the very few PLP-dependent enzymes that can be found in all living organisms (Singh et al., 2013). In animal cells, SHMT activity generates about 70% of the C1 groups (carbon units) needed for biosynthesis of thymidylate, purines and methione (Stover and Schirch, 1991). SHMT was first purified to homogeneity in 1963 (Schirch, 1982) and, from that date on, a lot of studies were carried out on this enzyme to understand the structure and its function. The interest on this topic is derived from the exceptionally broad substrate reaction selectivity that the enzyme shows in vitro.

In fact, in the presence of the appropriate substrate analogues, SHMT is able to catalyse transamination, racemisation, decarboxylation, aldol addition, and retroaldol reactions (Matthews and Drummond, 1990).

The increasing availability of crystal structures of the enzyme from various prokaryotic and eukaryotic sources contributed to clarify a number of observations previously acquired with classical biochemical studies. Up to now, SHMT crystal structures from three eukaryotic sources and three from prokaryotic sources are available: human liver cytosolic recombinant SHMT (hcSHMT) (Renwick et al., 1998), rabbit liver cytosolic recombinant SHMT (rcSHMT) (Scarsdale et al., 1999), murine cytoplasmatic SHMT (mcSHMT) (Doletha et al., 2000), *Escherichia coli* SHMT (eSHMT) (Scarsdale et al., 2000), *Bacillus*

staerothermophilus (bsSHMT) (Trivedi et al., 2002) and *Plasmodium vivax* (PvSHMT) (Chitnumsub et al., 2014).

In general, the fold type I group includes many of the best-characterized PLP-dependent enzymes. As is common to the other members of this group, each SHMT subunit, that associates into dimers in prokaryotes and tetramers in eukaryotes, folds into two domains: the central feature of the N-terminal, larger domain is a seven-stranded β-sheet, while the small, C-terminal domain is made by a three-stranded β-sheet, covered with helices on one side. The active site is located at the interface of the domains and is delimited by amino acid residues contributed by both subunits of the dimer. The crystallographic data greatly facilitated targeted mutagenesis and mechanistic studies. The characterization of the structural and functional properties of active site-mutants has solved some key questions regarding the catalytic mechanism of SHMT and its broad reaction specificity.

1.3.1 SHMT catalytic mechanism, details

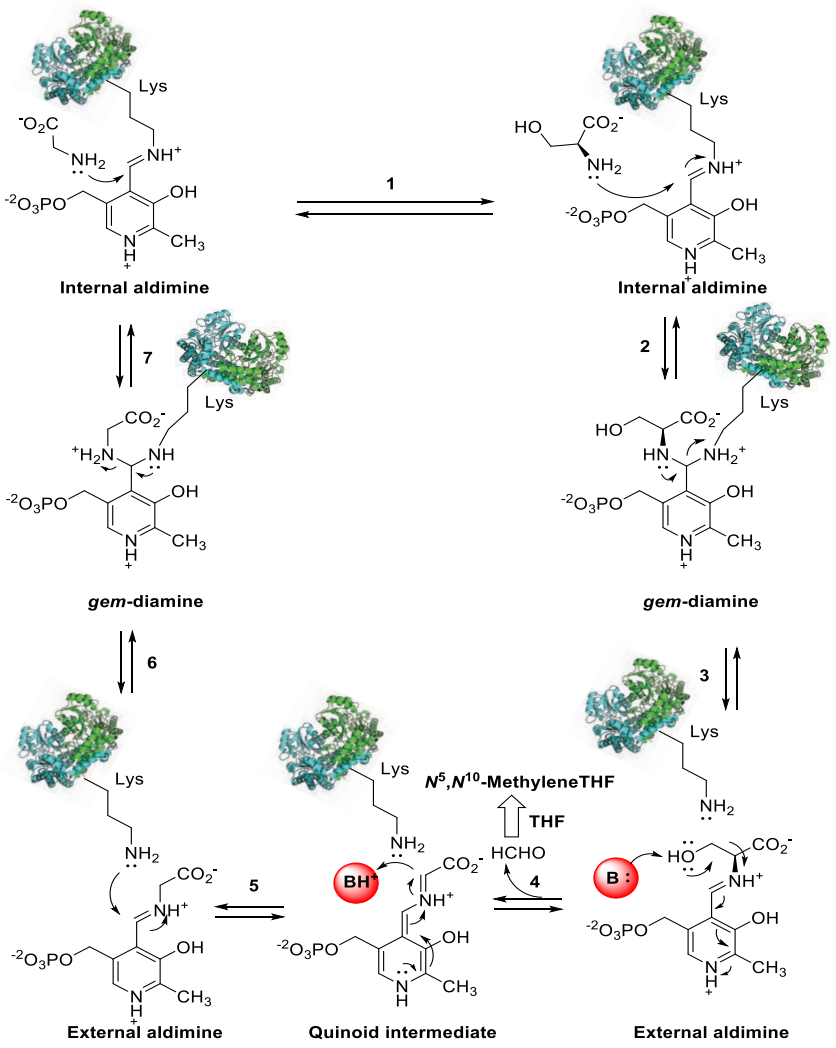
The catalytic mechanism of SHMT for the reversible retroaldol reaction consists of two co-ordinately catalysed reactions (Szebenyi et al., 2004) (Schirch and Szebenyi, 2005): cleavage of PLP-serine external aldimine to form PLP-glycine external aldimine and undissociated formaldehyde; followed by addition of formaldehyde to H4PteGlun (THF complex with the enzyme) to form N^5,N^{10}-MethyleneTHF. The serine and glycine external aldimines are rapidly

converted to the internal aldimine in which C4' of PLP is bound to the ε-amino group of Lys230. The reaction is completely reversible.

The first reaction step (**Scheme 2**) involves conversion of the internal aldimine, upon addition of L-serine, to L-serine external aldimine, via geminal diamine. The external aldimine is subsequently converted to a quinonoid intermediate, as a result of a direct attack of *N*5 of H_4PteGlu on Cβ of L-serine, resulting in C-Cβ bond cleavage to form glycine quinoid intermediate; H_4PteGlu is converted to CH_2-H_4PteGlu (i.e. N^5,N^{10}-MethyleneTHF). The product quinonoid intermediate is converted to Gly external aldimine, then to geminal aldimine and, finally, upon release of product Gly, the –CHO group of PLP interacts with $–NH_2$ group of the catalytic lysine (230 in the case of the $SHMT_{Sth}$) to form the initial internal aldimine.

In the geminal diamine complex of SHMT from *Bacillus stearothermophilus*, the orientation of PLP is changed by 16° from that in the internal aldimine. Upon conversion to external aldimine there is a further change in the orientation of 9° corresponding to a total change of 25° in the quinonoid intermediate from that in the internal aldimine (Bhavani et al., 2005). The orientation of the PLP changes slightly, as the amino acid substrate is released, and a covalent bond is formed between the cofactor and the lysine residue. The changing in the orientation of the PLP cofactor is a crucial feature for catalysis.

Different orientations of the PLP will be also observed in the structures described in this work.



Scheme 2: SHMT catalytic mechanism

The unligated enzyme exists as an internal aldimine, in which an active site lysine forms a Schiff base with the C4 of PLP. Upon reaction with L-serine, the internal aldimine is converted to L-serine external aldimine, via geminal diamine intermediate, in which the amino group of L-serine is bound to PLP through a Schiff base linkage. Then, the attack of *N*5 of H4PteGlu on Cβ of L-serine of the external aldimine generates the quinonoid intermediate and the formaldehyde is released. Finally, protonation of the quinonoid intermediate at Cα, mediated by the lysine, produces the internal aldimine, through the release of glycine.

With L-serine substrate, SHMT does not catalyse alternative reactions with folate-dependent mechanism. The currently accepted model attributes this reaction specificity to the existence of an “open” and a “closed” active site conformation. The substrates trigger the closed conformation, which precludes alternative reaction paths. In fact, in the absence of L-serine the enzyme active site is “open”, whereas in the presence of substrate the enzyme is in the “closed” form. In this form there is no access of solvent to the active site. The requirement for this conformational change is to increase substrate and reaction specificity, blocking side reactions. (Schirch et al., 1991).

1.4 SHMT industrial and pharmaceutical applications

1.4.1. SHMT as a biocatalysist

Aldolases are an interesting class of lyases that have received much attention due to their ability to catalyse asymmetric carbon-carbon coupling reactions, where their chiral-induction properties lead to stereo-chemically pure products, even when the starting materials are not-chiral substrates (Clapes, 2015) (Clapes, 2016).

Among them, the glycine-dependent aldolases (Liu et al., 2000) are interesting catalysts for the synthesis of β-hydroxy-α-amino acids.

β-Hydroxy-α-amino acids are an important class of natural products with biological activity on their own and also as constituents of many naturally occurring complex compounds, such as immunosuppressant (Yadav et al., 2005). They also constitute excellent intermediates for the synthesis of biologically relevant compounds, such as idulonic acid

mimetics (Miura and Kajimoto, 2001), the immunosuppressive lipid mycistericin D (Miura and Kajimoto, 2001) acyclic sugar analogues (Hiracuma et al., 1995) among others. Furthermore, these polyfunctional compounds might be useful building blocks for peptidomimetics and others non-proteinogenic peptide-like structures of biological interest. Despite their widespread applications, the synthesis of these compounds remains a challenging task. Their chemical synthesis, in fact, requires multiple steps and a very strict control of stereo-selectivity.

Moreover, acetaldehyde is known to be an important flavour component of several diary products. In milk fermentation for yoghurt manufacture, the amount of acetaldehyde produced by lactic acid bacteria (LAB) is determinant in the final flavour of the product (Xanthopoulos et al., 2001). *Streptococcus thermophilus* and *Lactobacillus bulgari*cus are reported to be responsible for the production of acetaldehyde in yoghurt fermentation (Wilkins et al., 1986). Although the major pathway in LAB for acetaldehyde production is currently unclear, it is reported that the serine hydroxymethyl transferase *S. thermophilus* (SHMT; EC 2.1.2.1) has threonine aldolase activity through a folate independent mechanism. (Chavez et al., 2002).

1.4.2. SHMT as a drug target

SHMT, together with dihydrofolate reductase (DHFR) and thymidylate synthase are essential for the synthesis of thymidylate, a unique metabolite required for DNA biosynthesis not obtained by the salvage pathway. The conversion of serine to glycine, catalysed by SHMT,

donates a one-carbon unit to tetrahydrofolate to produce N^5,N^{10}-MethyleneTHF (CH_2-THF). CH_2-THF is used in thymidine synthesis and is a precursor of other folate species that contribute to purine synthesis. Folates can also allow the regeneration of methionine from homocysteine and thus facilitate the generation of *S*-adenosylmethionine (SAM), the methyl donor for both DNA and histone methylation reactions that influence the epigenetic control of gene expression.

The alteration of cellular metabolism has been recently recognized as a hallmark of cancer. (Hanahan and Weinberg, 2011). Central to the metabolic reprogramming of cancer cells are the complex pathways involving folates, providing the essential precursors to sustain cancer cell growth and affecting cellular antioxidative and methylation capacities, thus supporting tumor homeostasis. (Schulze and Harris, 2012). Serine hydroxymethyltransferase (SHMT) is a key protein in this scenario.

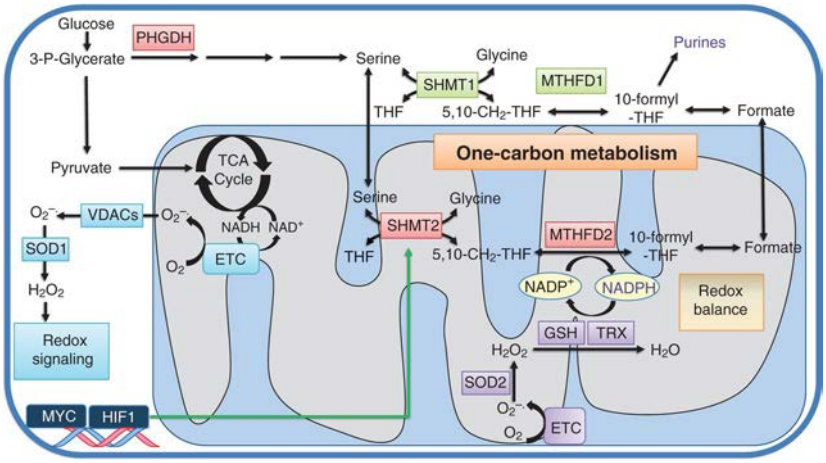


Figure 7: Serine metabolic derivatives with particular focus on one-carbon metabolism

Serine is a precursor of the non-essential amino acids glycine and cysteine. Glycine is in turn a precursor of porphyrins and is also incorporated directly into purine nucleotide bases and into glutathione (GSH). Serine is necessary for the production of sphingolipids via the synthesis of sphingosine, and serine is a headgroup, or headgroup precursor, for phospholipids. Additionally, serine supplies carbon to the one-carbon pool, which is involved in folate metabolism.

In the human genome, two SHMT genes are found; SHMT1, encoding the cytoplasmic isozyme (SHMT1), and SHMT2, encoding the mitochondrial one (SHMT2). (Garrow et al., 1993). The SHMT2 isozyme appears to be preferentially involved in the synthesis of mitochondrial thymidine monophosphate (dTMP), however, its chief function is probably to generate one-carbon units from serine, which are exported as formate into the cytosol and support one-carbon metabolism. (Fu et al., 2001) (Stover and Field, 2011). Production of glycine inside the mitochondria, required for heme biosynthesis and further production of one-carbon units by the glycine cleavage system, is also a pivotal function of SHMT2 (Kao et al., 1969) (di Salvo et al.,

2013b). SHMT1 participates as scaffold proteins to the synthesis of nuclear dTMP, undergoing nuclear localization in the S-phase during the thymidylate cycle along with thymidylate synthase (TS) and dihydrofolate reductase (DHFR) in order to limit the uracil misincorporation in DNA. (Anderson et al., 2012) (MacFarlane et al., 2011).

SHMT2 has been shown to be a crucial factor in the serine/glycine metabolism of several cancer cell types, including colon and breast, and thus represents a hot target for selective anticancer drugs. (Jain et al., 2012). Lung cancer is the second most common cancer expected to occur in men and women in 2014 and the first leading cause of cancer-associated deaths worldwide. (Siegel et al., 2014) In 2014 Zhang *et al.* (Zhang et al., 2012) demonstrated that SHMT1 is overexpressed in tumor-initiating lung cancer cells, suggesting that the function of this gene could be cancer cell type specific.

Serine hydroxymethyltransferase has also been studied as a potential target for anti-malarian chemotherapy (Chitnumsub et al., 2014). The expression of the Plasmodium SHMT gene is noticeably increased during late trophozoite to schizont stages when high levels of folate and nucleotides are needed for cell multiplication process, emphasizing the indispensable role of this enzyme (Nirmalan et al., 2002). Therefore, further investigation into the mechanism of *Plasmodium* SHMTs inhibition is of interest such that the possibility of developing specific inhibitors against the enzyme.

1.5 Serine hydroxymethyltransferase from *Streptococcus thermophilus*

In previous works in collaboration with the chemical engineering group at the Universitat Autonoma de Barcelona (UAB), the *glyA* gene encoding an SHMT enzyme with L-threonine aldolase activity (LTA) was isolated from *Streptococcus thermophilus* YKA-184 chromosomal DNA (Vidal et al., 2005)

LTA activity consists in the reversible cleavage of L-threonine into glycine and acetaldehyde (**Fig.8**).

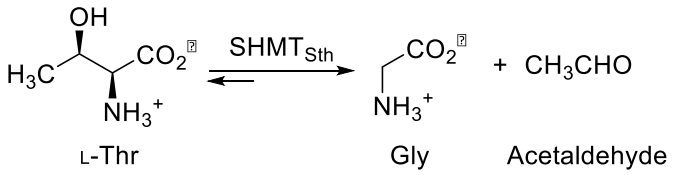
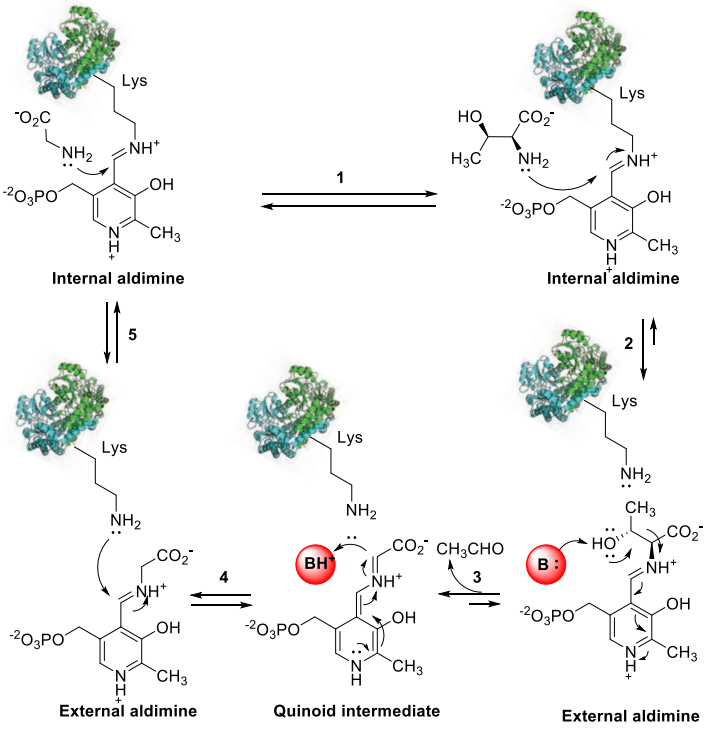


Figure 8: LTA activity shown by $SHMT_{Sth}$

$SHMT_{Sth}$ catalyses the reversible cleavage of L-threonine to glycine and acetaldehyde.

Initially, the cofactor PLP is covalently bound, as a Schiff base, to the side chain of Lys230. The addition of L-threonine substrate to $SHMT_{Sth}$ results in the formation of external aldimine complex. The retro-aldol cleavage involves deprotonation of the L-threonine external aldimine by an, up to now, unidentified base, leading to the formation of glycine external aldimine intermediate. The enzyme acts as an electron sink, storing electrons from cleaved substrate bonds and dispensing them for the formation of new linkages with incoming protons or second substrates.



Scheme 3. SHMT$_{Sth}$ retro-aldol cleavage scheme reaction

The hypothesized retro-aldol cleavage mechanism catalysed by LTAs is represented in the scheme. The deprotonation of L-thr C is carried out by an unidentified base species, represented in red in the scheme, that triggers the quinoid intermediate formation.

LTAs can classified as specific threonine aldolases (TA), low-specific TA or specific *allo*-TA, depending on their stereo-specificity towards the β-carbon of threonine (Vidal et al., 2005).

It was observed that $SHMT_{Sth}$ K_M for L-allo-threonine was 38-fold higher that for L-threonine, suggesting this enzyme can be classified as a specific L-*allo*-threonine aldolase.

It has long been argued that both SHMT and LTA activities are catalysed by the same protein. However, several studies have demonstrated that, in some organism, are discrete enzymes, for example rat liver SHMT does not exhibit TA activity (Ogawa and Fujioka, 2001). In *E. coli*, both reactions depend on one single protein (Plamann et al., 1983) (Schirch et al., 1985). Until now, only one protein with both LTA and SHMT activities has been reported in *S. thermophilus* (Chavez et al., 2002).

When tested for aldol addition reactions with non-natural aldehydes, such as benzyloxyacetaldehyde and (*R*)-*N*-Cbz-alaninal, two possible β-hydroxy-α-amino acid diasteroisomers were produced, with high stereoselectivity at the amino group and moderate at the β-hydroxyl group (Gutierrez et al., 2008).

The preliminary results obtained (Gutierrez et al., 2008) showed that this enzyme was a promising biocatalyst for the stereoselective synthesis of β-hydroxy-α-amino acids, therefore mutational studies could improve the catalytic properties of the enzyme. Particularly we are interested in expanding the nucleophile selectivity of SHMT towards the addition of other amino acids rather than glycine. This will constitute a new methodology for the construction of stereogenic tetrasubstituted carbon centres and will give access to the synthesis of chirally pure α,α-dialkylated β-hydroxy-α-amino acids.

Expanding the nucleophile selectivity in aldolases is as essential as challenging. Whereas most aldolase-type enzymes tolerate a broad variety of non-natural aldehyde electrophiles with good catalytic rates,

they generally share high substrate specificity for their nucleophile, as reiteratively documented for a broad variety of distinct aldolases from various sources (Brovetto et al., 2011, Windle et al., 2014). Even small structural variations in the nucleophile (Chen et al., 1992; Fessner *et al.*, 1994; Arth *et al.*, 1998; Watts *et al.*, 2004) with specific aldolases resulted in a decrease of activity of up to several orders of magnitude, reflecting the strong influence of steric and electronic factors on the intricate binding environment required to stabilize the highly ordered bisubstrate transition state.

Based on the sequence alignment studies of the $SHMT_{Sth}$ with α–methylserine hydroxymethyl transferase (MSHMT) from *Paracoccus sp.*, three residues were identified that could be involved in the donor selectivity (Hernandez et al., 2015a) Y55, Y65 and H229. Hence, the $SHMT_{Sth}$ variants Y55T, Y65H, H229T were constructed and tested as catalyst for the aldol addition of glycine, L-serine, D-serine, L-threonine and D-threonine to (benzyloxy)acetaldehyde. Very positive results were obtained with the variants in position 55, and then a focused library of mutants (94 clones) was generated by site direct mutagenesis at residue Y55 and the catalytic activity in the presence of different reaction substrates, using (benzyloxy)acetaldehyde as acceptor, was tested. Under identical reaction conditions, Y55T/C/S variants gave remarkable results.

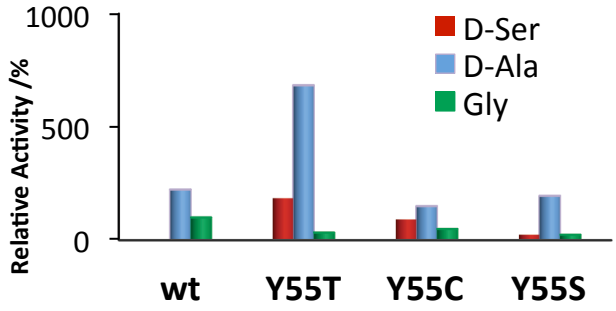
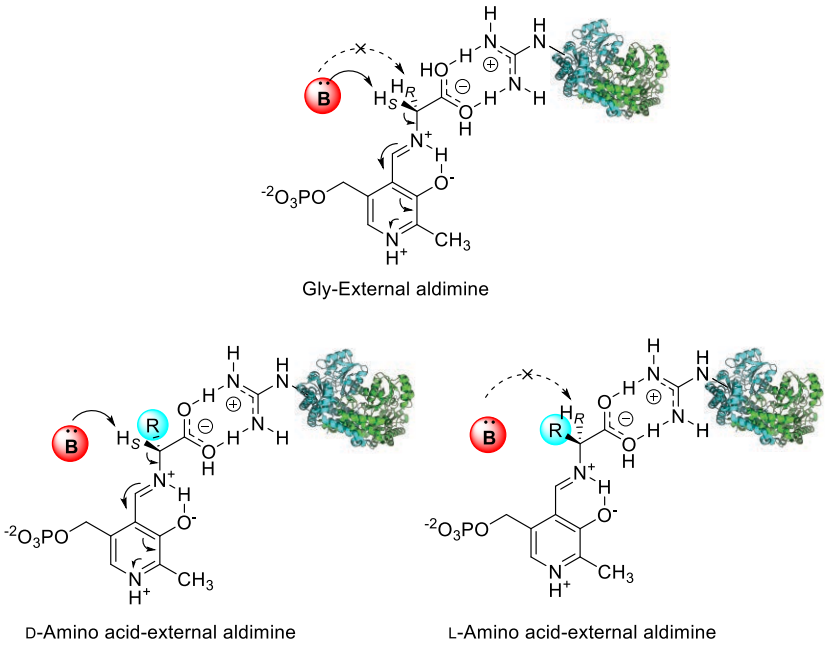


Figure 9: Enzymatic activity of $SHMT_{Sth}$ variants studied

Relative initial reaction rates in % ($V_0/V_0^{Gly\_SHMT\ wt}$) of the addiction of D-serine (in red), D-alanine (in blue) and glycine (in green) to (benzyloxy)acetaldehyde catalyzed by $SHMT_{Sth}$ variants and the wild-type.
V_0 =initial reaction rate; $V_0^{Gly\_SHMT\ wt}$ =initial reaction rate of the aldol addition of to for wild-type $SHMT_{Sth.}$ Data from (Hernandez et al., 2015a)

The best donor for $SHMT_{Sth}$ variants and the *wild-type* was D-alanine. $SHMT_{Sth}$ Y55T was the best catalyst for D-alanine and D-serine donors, whereas the *wild-type* enzyme was the best for glycine, it was active with D-alanine but was inactive with D-serine. No aldol reaction was detected with the corresponding L-amino acids. The reason lies in the catalytic mechanism of the SHMT. The common principle of all known aldolases concerning the retention of configuration makes the L-selective SHMT to abstract the pro-*S* proton during the activation of the amino acid nucleophile. This pro-*S* proton is located in a position accessible to the acid-base machinery. To activated other amino acids than glycine as nucleophiles, the α-proton must be accessible to the acid-base, and this only can be accomplish by using D-amino acid as nucleophilic components. In this case the equivalent to the pro-*S* proton of glycine is abstracted (**Scheme 4**).



Scheme 4: Orientation of the external aldimine from L and D amino acids in the SHMT active site

In the case of D-amino acids, the pro-*S* proton is located in a position accessible to the acid-base (in red) machinery. In the case of L-amino acids, the proton has an opposite orientation and it is inaccessible to the base for the abstraction.

The variants produced, Y55T, Y55S and Y55C, were inactive in the retroaldol reaction of L-threonine, contrarily to that observed with SHMT *wild type*. Some authors suggested that this might be due to the loss of stabilizing interactions caused by mutating the tyrosine to other amino acids (Bhavani et al., 2008). However this hypothesis does not fully explain the absence of retroaldol activity observed and up to now no clear explanation has emerged.

To get insight into the precise alteration of the active site as a consequence of Y55 modification is one of the goals of the present

work, which will shed some light into the intricate mechanism of the $SHMT_{Sth}$ To this end, structural studies of the *wild-type* protein from *Streptococcus thermophilus* and its most catalytically relevant variants are the topic of the work in this thesis.

Results and Discussion

1.6 $SHMT_{Sth}$: overexpression and purification of *wild-type* and variants

The plasmid pQE40-SHMT, containing the gene for expression of serine hydroxymethyl transferase from *Streptococcus thermophilus* ($SHMT_{Sth}$) was a generous gift from "Departament d'Enginyeria Química" of the Universitat Autònoma de Barcelona. It includes the $SHMT_{Sth}$ genetic sequence corresponding to the mature protein and incorporates a tag of 6 histidines at the N-terminus of the protein to facilitate purification by affinity chromatography. The protein was expressed in M-15 [pREP-4]. This strain is lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase. It also contains a plasmid, pLys-S, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG.

The $SHMT_{Sth}$ variants Y55T and Y55S were produced by site direct mutagenesis technique and tested in our research group for previous studies regarding the catalytic activity of mutated $SHMT_{Sth}$.

The optimal conditions for the solubilisation of *wild-type* and variants during cell lysis were established: 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, pH 8. After supernatant clearance by centrifugation, a two-step purification protocol follows. First the filtered supernatant is passed through a nickel-chelating column that traps histidine-tagged proteins, which are eluted by a gradient of imidazole. The purest eluting fractions are pooled, concentrated and injected to a gel filtration column. The elution fractions were pooled, concentrated and finally dialysed against HEPES 10 mM pH 7.5.

The final yield, after protein purification, was 20 mg of $SHMT_{Sth}$ per 1L of culture. The protein is obtained with a level of purity higher than 95% as assessed by SDS-PAGE (**Fig.10**) and its monodispersity in solution was evaluated by Dynamic Light Scattering technique (DLS).

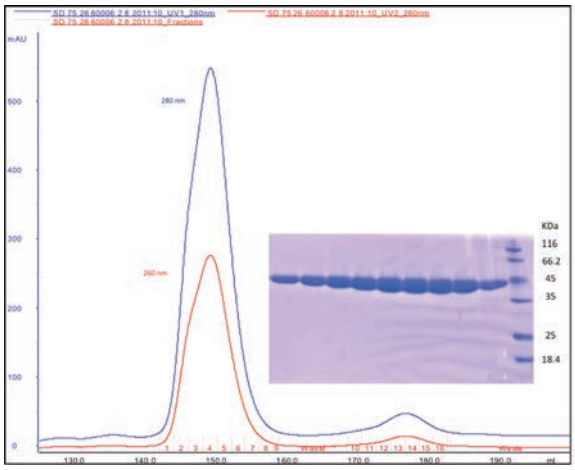


Figure 10: $SHMT_{Sth}$ purification chromatogram and SDS-gel.

$SHMT_{Sth}$ is purified with superdex 200/16-60 column (GE Healthcare). The absorbance at 280 and 260 nm are shown in blue and red lines respectively. The eluted fractions from the first peak are loaded into a poly-acrylamide denaturating gel (on the right hand side of the picture). As can be seen, a single pure band of the protein is obtained that migrates as a molecule of 45 kDa.

1.7 $SHMT_{Sth}$ *wild-type*

1.7.1 Crystallization and data collection of $SHMT_{Sth}$ *wild-type*

In order to elucidate the tridimensional structure of $SHMT_{Sth}$ *wild-type* in atomic detail, crystallization experiments were performed and suitable crystals for the X-ray diffraction were obtained in two different crystallization buffers. The monodisperse protein in 10 mM HEPES buffer was crystallized in its *apo*-form and in complex with the cofactor PLP and reaction substrates (L-threonine and glycine).

Monoclinic colourless crystals appeared after three days in a crystal buffer containing 0.8-1.2 M sodium citrate, 0.1 M cacodylate pH 6.5. In the absence of PLP cofactor, citrate displaces PLP, rendering colourless crystals of *apo*-$SHMT_{Sth}$ (**Fig.11/A**)

Tetragonal yellow crystals, corresponding to the $SHMT_{Sth}$ complexes (or *halo*-$SHMT_{Sth}$ structures) appeared after few days in crystal buffers containing 0.85 M sodium citrate, 1 M cacodylate pH 6.5, 0.1 mM PLP and 100 mM of L-threonine for L-threonine complex, and 1.1 M sodium citrate, 0.1 M cacodylate pH 6.5, 5 mM PLP and 17 mM glycine for glycine complex (**Fig.11/B-C**). The presence of PLP gives the characteristic yellow colour to these crystals. For the *apo*-SHMTSth a dataset, with a total of 200 images spanning 200°, was collected in frames of 1° rotation and with an exposure time of 2s, using a charged-coupled device (CCD) area detector and a fixed wavelength. The final resolution obtained was 2.3 Å. For the two $SHMT_{Sth}$ Gly and L-Thr complexes two datasets, with a total of 100 and 180 images spanning 90°, were collected in frames of 0.5° rotation and 1° rotation with an exposure time of 23s and 1s respectively, using a charged-coupled device (CCD) area detector and a fixed wavelength.

The final resolutions obtained were 2.4 and 2.0 Å. All the data were collected from the beamline ID29 at the ESRF synchrotron in Grenoble, France. We thank Igor Zelen, who performed these crystallization experiments.

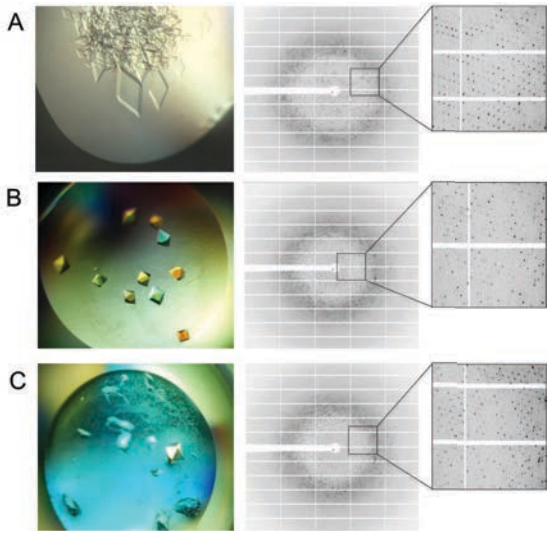


Figure 11: Crystals of *apo*-SHMT$_{Sth}$ and complexes and their corresponding X-ray diffraction patterns.

Images of the crystallization drops and X-ray diffraction patterns of typical crystals of the *apo*-SHMT$_{Sth}$ and complexes are shown. (**A**) Monoclinic colourless crystals corresponding to SHMT$_{Sth}$ apo-form and the best dataset diffracting to 2.3 Å; (**B**) Tetragonal yellow crystals corresponding to SHMT$_{Sth}$ in complex with PLP and glycine and the best dataset diffracting to 2.2 Å resolution; (**C**) Yellow tetragonal crystal of SHMT$_{Sth}$ in complex with PLP and L-threonine and the corresponding data set diffracting to 2.4 Å resolution.

1.7.2 SHMT$_{Sth}$ *wild-type* data processing, crystal structure solution and refinement

Diffraction data collected for the *wild-type* protein were indexed, integrated, scaled and reduced to unique reflections with XDS (Kabsch, 2010).

The crystals of $SHMT_{Sth}$ *wild-type* in its *apo*-form and the complexes belong to two different space groups: monoclinic *C*2 and tetragonal $P4_12_12$ respectively. Space group *C*2 with unit-cell parameters a=202.1 Å, b=113.4 Å, c=133.3 Å, α=γ=90.0, β=93.8, and space group $P4_12_12$ with unit-cell parameters *a*=115.4 Å, *b*=115.4 Å, *c*=191.4 Å, *α=β=γ*= 90.0 Å.

The calculated Matthews coefficient for WT*apo* was of 4.43 $Å^3$/Da, which is consistent with four copies of 43 KDa in the asymmetric unit (a.u.). The structure was solved by the molecular replacement method (Read, 2001) using Phaser from the CCP4 suite (McCoy et al., 2007), 2007) and employing the structure of SHMT from *Bacillus Stearothermophilus* as a search model (PDB: 1KKJ) (Trivedi et al., 2002). The model was manually completed using COOT (Emsley and Cowtan, 2004) and was refined with Refmac5 (Murshudov et al., 1997). The solution obtained was well discriminated from all others, presenting a Z-score for the translation search of 11.6 that, according to statistical criteria implemented in this program, indicates unequivocally that it is the correct one. Furthermore, the crystal packing derived from the solution presented no clashes between molecules of the a.u. and the ones related by symmetry demonstrating that the solution is physically sound. Accordingly, a visual inspection of the packing confirmed the correct distribution of the molecules within the crystal.

The calculated Matthews coefficients for WTgly and for WTLthr were of 2.59 and 2.32 $Å^3$/Da respectively, which is consistent with two copies in the asymmetric unit. The phase problem was solved by the molecular replacement method (Read R. J. et al., 2007) using Phaser through the CCP4 suite (McCoy et al., 2007) and employing the

WT*apo* structure as a search model. The solution obtained presented a Z-score of 12. Also in these cases the solution presented no clashes between molecules of the asymmetric unit.

Diffraction data and refinement statistics of the $SHMT_{Sth}$ *apo*-form and of binary complexes obtained from are summarised in **Table_1**.

The values of the mean global indicators in refinement of these three datasets, R-free (between 0.20 and 0.22) and R-work (between 0.15 and 0.19), are in line with those of good quality data sets between 2.05 and 2.4 Å resolution (Kleywegt and Brunger, 1996). Atomic contacts and structure validation were checked using Molprobity server (Davis et al., 2007)

The asymmetric unit contains four molecules (A, B, C and D) in the case of WT*apo* and two (A and C), molecules in the case of WTgly and WTLthr. In all the molecules in both *apo*- and *halo* forms, the structure was traced from residue 7 to residue 404, with the exception of residue 110 to 119 for the chains B and C of WT*apo* structure, as explained later.

For convenience, the structure of $SHMT_{Sth}$ *wild-type apo*-form will be indicated, from now on, as "**WT*apo***"; the structure of $SHMT_{Sth}$ *wild-type* in complex with glycine will be indicated as "**WTgly**"; and the structure of $SHMT_{Sth}$ *wild-type* in complex with L-threonine will be indicated as "**WTLThr"**.

	WTapo	**WTgly**	**WTLthr**
Data collection			
Source	ID23-EH1	ID214-EH1	ID29
Wavelenght (Å)	0.9699	0.9330	1.0000
Space group	*C*2	$P4_12_12$	$P4_12_12$
Cell dimensions			
a, *b*, *c* (Å)	201.9-113.7-132.9	115.4-115.4-191.5	115.2-115.2-191.1
a, *β*, *γ* (°)	90.0-93.5-90.0	90.0-90.0-90.0	90.0-90.0-90.0
A.U. content	two dimers	dimer	dimer
Resolution (Å)	83.3–2.0 (2.4–2.3)	99.0-2.4 (2.1-2.0)	99.0-2.0 (2.5-2.4)
Unique reflections	133216 (15884)	186803 (24758)	51708 (6122)
Completeness (%)	99.8 (99.8)	99.8 (99.0)	99.8 (99.8)
Redundancy	3.6 (3.6)	8.1 (8.0)	14.1 (10.0)
I/σI	10.3 (2.2)	14.4 (2.2)	16.6 (2.4)
Refinement			
R_{work} / R_{free}	0.19/ 0.22	0.15/0.20	0.17/0.20
No. atoms			
Protein	12370	6278	6281
Ligand	96	69	52
Water	578	392	354
B factors			
Protein	43.30	43.60	31.20
Ligand/ion	85.5		
R.m.s. deviations			
Bond lengths (Å)	0.012	0.009	0.010
Bond angles (°)	1.18	1.22	1.28
PDB code	4WXB	4WXG	4WXF

Table 1: Crystallographic statistics on data collection and refinement
Values within parentheses correspond to the last resolution shell.

1.8 $SHMT_{Sth}$ Y55S and Y55T variants

1.8.1 Crystallization and data collection of $SHMT_{Sth}$ Y55S and Y55T variants

To characterize the enzyme catalytic pocket in its atomic details and to define the role of each amino acid present in it, the X-ray structures of $SHMT_{Sth}$ Y55S and Y55T variants, in complex with PLP and relevant reaction substrates, were determined. Crystallization experiments were performed and suitable crystals for the X-ray diffraction were obtained in crystallization conditions similar to the ones used for the *wild-type* protein with the addition of the cofactor and three different substrates: L- and D-threonine and D-serine.

The Y55S variant, co-crystallized with D-threonine, renders monoclinic colourless crystals after a few days in a crystal buffer containing 1.2 M sodium citrate, 0.1 M cacodylate pH 6.5, 0.1 mM PLP and 5 mM D-threonine. The absence of PLP cofactor in the active site of the protein (further explained later) renders colourless protein crystals **(Fig.12A)**.

The Y55S variant co-crystallized with L-threonine and D-serine, renders tetragonal yellow crystals after a few days in a crystal buffers containing 1.2 M sodium citrate, 0.1 M cacodylate pH 6.5, 0.1 mM PLP and 5 mM L-threonine or 0,1 mM PLP and 5-10 mM D-serine respectively (**Fig.12/B-C-D**).

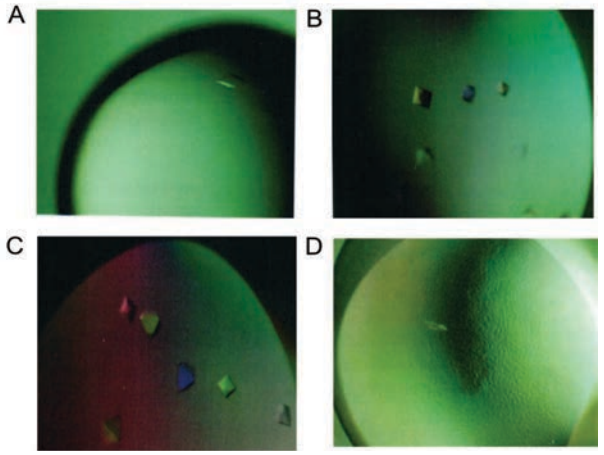


Figure 12: Crystals of $SHMT_{Sth}$ variants in complex with substrates

Images of the crystallization drops of the $SHMT_{Sth}$ Y55S variant in complex with (**A**) D-threonine; (**B**) L-threonine; (**C**) D-serine 5mM; (**D**) D-serine 10mM. In the case of $SHMT_{Sth}$ Y55S variant crystallized with 5mM of D-serine just one crystal was found in the crystallization drop.
The corresponding diffraction images are not shown here, because comparable in aspect and quality to the ones collected for the *wild-type* protein and they are present in the "Experimental procedure" session of this thesis.

The Y55T variant, co-crystallized with D-serine, rendered thin crystals not suitable for X-ray diffraction experiments. Micro-seeding experiments were necessary to improve crystal quality: the thin crystals obtained were used as "seeds" for the crystal nucleation and growth of larger protein crystals, as shown in **Fig.13**.

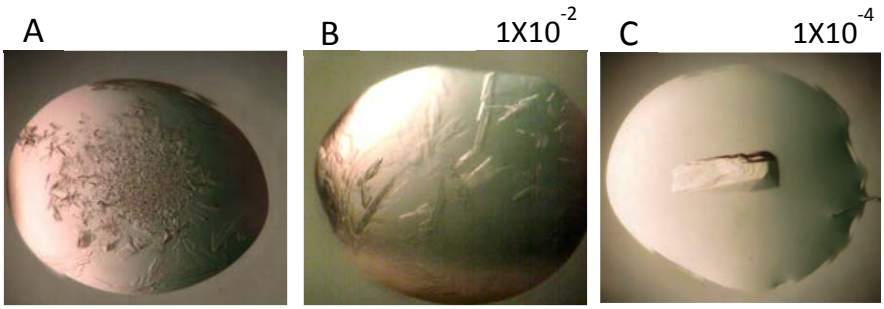


Figure 13: Crystals of $SHMT_{Sth}$ Y55T variant in complex with D-serine before and after micro-seeding experiments.

(**A**) Crystallization drop, corresponding to the $SHMT_{Sth}$ Y55T variant, co-crystallized with D-Ser. Many thin crystals, impossible to separate, are present. (**B**) Crystallization drop obtained after micro-seeding experiment, using a dilution of seeds equal to $1x10^{2}$. The thin crystals were removed from the drop in (A) and placed in a microcentrifuge tube containing sead bead (Hampton Research) and a mother liquor (100X). The improvement in crystal appearance is evident, but the crystals are steel thin. (**C**) A dilution equal to $1x10^{-4}$ allows growing a larger, single crystal, easy to extract from the mother liquor and suitable for X-ray diffraction experiments.

For the complexes of the $SHMT_{Sth}$ variants, five datasets with a total of 3600 images spanning a full rotation each were collected in frames of 0.25° rotation and with an exposure time of 0.25 min, using a PILATUS 6M detector and a fixed wavelength. The datasets were collected from the beamline XALOC at the ALBA synchrotron in Barcelona, Spain.

1.8.2 Data processing, crystal structure solution and refinement of $SHMT_{Sth}$ Y55S and Y55T variants

Diffraction data collected for the variants, as in the case of *wild-type* protein data, were indexed, integrated, scaled and reduced to unique reflections with XDS (Kabsch, 2010)

The crystals of Y55S variant in complex with PLP and D-threonine belong to *C*2 space group, with unit cell parameters *a*=200.3 Å, *b*=113.6 Å, *c*=133.1 Å, $\alpha=\gamma=90.0$ $\beta=94.5$.

The crystals of Y55S variant in complex with PLP and L-threonine belong to $P4_12_12$ space group, with unit cell parameters *a*=114.7 Å, *b*=114.7 Å, *c*=92 Å, $\alpha=\beta=\gamma=90.0$.

For the Y55S variant crystallized in complex with PLP and D-serine, two datasets corresponding to both crystal forms were collected. One crystal belonged to the $P4_12_12$ space groups with unit-cell parameters *a*=114.7 Å, *b*=114.7 Å, *c*=192.8 Å, $\alpha=\beta=\gamma=90.0$ 94.1, and another crystal belonged to space group *C*2, with unit cell parameters *a*=114.7 Å, *b*=113.6 Å, *c*=132.4 Å, $\alpha=\gamma=90.0$ $\beta=94.1$.

And, finally, the crystal of Y55T variant in complex with PLP and with D-serine belonging also to space group *C*2, showing unit-cell parameters a=200.7 Å, b=112.9 Å, c=131.9Å, α= γ=90.000, β=93.1 was determined.

The diffraction quality of the tetragonal crystals was comparable, yielding resolution limits between 1.9 and 2 Å; lower resolution values, 2.3 Å and 2.4 Å were achieved for the complexes in the monoclinic space group. Diffraction data and refinement structural statistics for the structures of the variants are summarised in **Table 2**.

The values of both R-free and R-work (0.17 and 0.22) for all the structures are reasonable for data sets between to 2.4 and 2.05 Å resolution (Kleywegt and Brunger, 1996).

The structures belonging to the *C*2 space group contain four molecules in the asymmetric unit (chain A, B, C, D), the tetragonal ones just two molecules (A, C), corresponding to a dimer, that is the oligomeric state for the catalytically active species.

For convenience, the structure of Y55S $SHMT_{Sth}$ in complex with L-threonine will be indicated as "**Y55SLThr**", and in complex with D-threonine will be indicated as "**Y55SDThr**"; the structure of Y55S $SHMT_{Sth}$ in complex with PLP and D-serine crystallised in $P4_12_12$ space group will be indicated as "**Y55SPLS**" and in complex with PLP and D-serine crystallised in *C*2 space group will be indicated as "**Y55SDSerPLP**". Finally, the structure of Y55T in complex with L-threonine will be indicated as "**Y55TPLS/DSerPLP**".

	Y55SLThr	**Y55SDThr**	**Y55SPLS**	**Y55SDSer/PLP**	**Y55TPLSDSerPLP**
Data collection					
Source	XALOC	XALOC	XALOC	XALOC	XALOC
Wavelength (Å)	0.9699	0.9330	1.0000	0.9697	0.9798
Space group	$P4_12_12$	$C2$	$P4_12_12$	$C2$	$C2$
Cell dimensions					
a, *b*, *c* (Å)	114.7-114.7-192	200.3-113.6-133.1	114.7-114.7-192	114.7-113.6-132.4	200.7-112.9-131.9
a, *β*, *γ* (°)	90.0-90.0-90.0	90.0-94.5-90.0	90.0-90.0-90.0	90.0-94.1-90.0	90.0-93.1-90.0
A.U. content		two dimers	dimer	two dimers	two dimers
Resolution (Å)	98.4–1.97 (2.0-2.0)	99.0-2.5 (2.5-2.1)	99.9-1.47 (2.0-1.9)	98.0-2.4(2.4-2.3)	98.0-2.2(2.2-2.1)
Unique reflections	91521(82641)	106614(83565)	215988(11437)	80334(956)	165927(21511)
Completeness (%)	99.9 (99.8)	98.9 (99.0)	99.9 (98.0)	96.1(63.4)	99.3(97.8)
Redundancy	3.6 (3.6)	8.1 (8.0)	2.98(0.27)	1.59(0.08)	6.5(6.7)
R_{merge}	1.2-0.06	1.6-0.09	1.75-0.07	0.35-0.05	0.58-0.05
I/σI	10.3 (2.2)	9.8 (1.6)	9.16(0.75)	13.46(1.39)	17.9(2.03)
Refinement statistics					
Resolution (Å)	98.48 – 2.0	132.7 - 2.14	98.63 – 1.93	132.13 – 2.33	47.71-2.12
No. reflections	87017	161543	92342	76366	157619
R_{work} / R_{free}	0.17/0.17	0.17/0.17	0.16/0.16	0.17/0.22	0.23/0.28
No. atoms					
Protein	6246	12492	6234	12476	12503
Ligand/ion	53	51	62	79	80
Water	392	757	503	386	396
B factors					
Protein	49.15	43.01	45.04	43.45	59.88
Ligand/ion	49.95	52.60	51.18	36.63	74.10
Water	52.02	46.83	48.81	33.09	56.71
R.m.s. deviations					
Bond lengths (Å)	0.019	0.0204	0.0185	0.0145	0.0184
Bond angles (°)	1.960	2.137	1.864	1.773	1.940

Table 2: Crystallographic statistics on data collection and refinement

Values within parentheses correspond to the last resolution shell.

1.9 $SHMT_{Sth}$ *wild-type* structures

1.9.1 $SHMT_{Sth}$ *wild-type apo*-form (WT*apo*) overall structure

The final model of $SHMT_{Sth}$ contains four equivalent (they show an r.m.s.d. of 0.20 over the 410 Cα carbons, when superposed) 45-kDa chains (A-B-C-D) of 410 residues each (7-417), four molecules of citric acid, four molecules of cacodylate (one bounded in each of the four active sites), and 578 water molecules in the asymmetric unit. No electron density corresponding to residues 1-6 and the attached N-terminal His-tag can be seen, presumably due to disorder; whereas, the remaining polypeptide chain is well defined in the electron density map and only eight residues (122-131) in chain B and C, contributing to the catalytic pocket as explained later in this session, appear to be disordered and, consequently, are not included in the final model.

The stereochemistry of the refined model is good; the Ramachandran plot shows that 96% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt and Jones, 1996) and only nineteen non-glycine, non terminal amino acid residues lie outside the core or most restrictive extension region of allowed values. Three outliers regard the catalytic lysines 230 in three of the four chains (A, B, D). This can be explained considering the flexibility required by the active site to undergo changes during catalysis and stabilize the reaction intermediates. Four outliers regard the Thr-353 in all the chains, fact that could be explained considering its close proximity to the catalytic lid of the active site that is a very flexible part of chain. The Ramachandran plots, corresponding to each structure are illustrated in the “Experimental procedures” section.

Analysis of the WT*apo* structure reveals that the monomer fold is similar to other PLP-dependent enzymes of the α-class, such as SHMT

from *Bacillus Stearothermophilus* (*bs*SHMT), with an r.m.s.d. of 0.57, but exists as a dimer in the crystalline state (**Fig.14**). The crystallographic asymmetric unit, in fact, consists of two tight dimers, stabilized by a large number of hydrogen bond and van der Waals interactions and many bound solvent molecules, and no tetrameric contacts are visible in the crystal structure. This is consistent with the gel filtration experiments, which shows that the protein elutes at a volume corresponding to a dimer. (**Fig.10**). The contact surface area of $SHMT_{Sth}$ monomers in the dimer is 17883 $Å^2$.

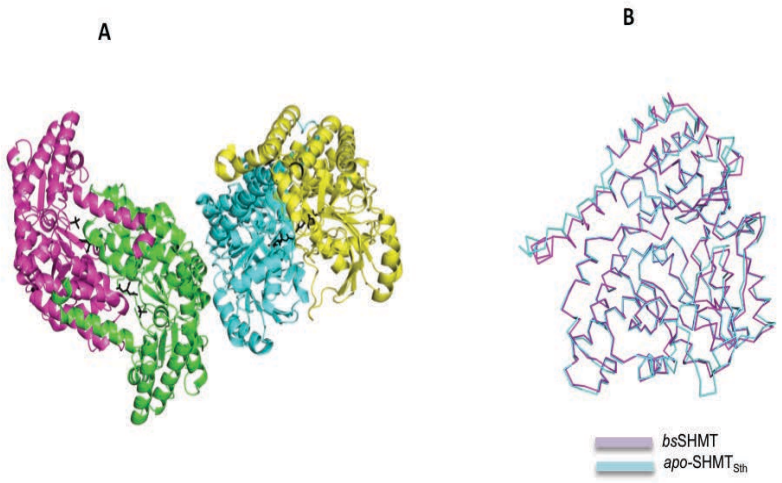


Figure 14: Quaternary structure of WT*apo* in comparison with SHMT from *Bacillus stearothermophilus*

(A) Cartoon representation of the $SHMT_{Sth}$ WT*apo* structure. Subunit A is coloured in green, subunit B in cyan, subunit C in magenta and subunit D in yellow. A-D and B-C form the tight dimers, which are related by a horizontal twofold axis. Cacodylate and citric acid, both present in the active site, are shown in ball-and-stick representation in black. **(B)** Superposition of $SHMT_{Sth}$ WT*apo* (cyan) and *bs*SHMT (magenta) monomers reveals a very similar overall organization.

All the SHMTs studied up to now are homodimers or homotetramers. The mammalian SHMT examined so far have been found to be tetrameric. In contrast, the *E. coli* SHMT has been described as both tetramer and dimer (Scarsdale et al., 2000), depending on the expression level in *E. coli*. And the other SHMTs published up to now, coming from prokaryotic sources (*e*SHMT, *bs*HMT, *mtb*SHMT), have been found as dimers (Scarsdale et al., 2000) (Trivedi et al., 2002). For protein activity, the dimer appears to be the minimum structure required, being the active site present at the interposition between two monomers, as further explained later.

The monomer fold can be described in terms of three domains (**Fig.15**): the N-terminus, the "large" domain and the "small" domain. The N-terminus (residues 7-284 in the structure) mediates inter-subunit contacts and it can be further divided in two sub-domains: "large" and "small" domains. The "small" domain (residues 7-84) folds into four helices and three β strand. The "large" domain (residue 85-284) binds the cofactor and folds into an αβα structure, consisting of a seven-stranded mixed β-sheet flanked by α-helices on both sides. The sheet is flanked by six helices on one side and by one helix on the other. Finally, the C-terminal "small domain" (residues 285-411) folds into an αβ sandwich. This sheet packs on one side against the large domain and is shielded from the solvent by two helices on the other side.

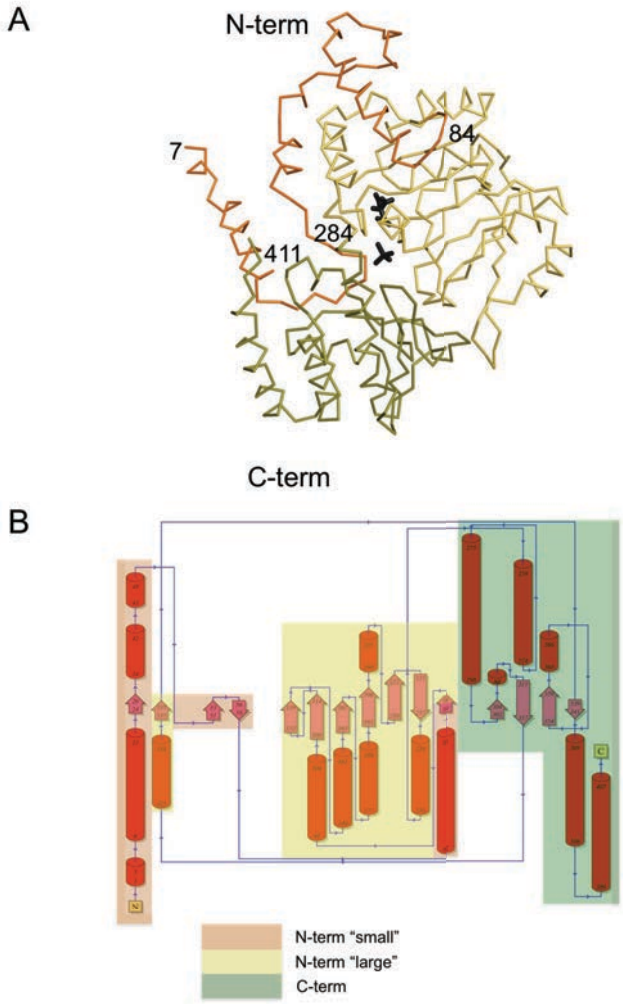


Figure 15: The overall fold of WT*apo* monomer

(**A**) Ribbon diagram of the monomer fold. The small N-terminal domain is shown in orange, the large N-terminal domain in yellow and the C-terminal domain in green. The PLP cofactor is shown in black in ball-and-stick representation. **(B)** The topology scheme of the $SHMT_{Sth}$ monomer is shown here, representing the helices as rods, β-strands as arrows.

1.9.1.1 $SHMT_{Sth}$ *wild-type apo*-form catalytic site architecture

The active site is located at the interface between two monomers, thus amino acids coming from different chains contribute to one catalytic site. Inside each active site of WT*apo* a molecule of cacodylate and a molecule of citric acid are present, interacting with different residues. Both species come from the buffer employed for the crystallization (**Fig.16**).

As demonstrated for other proteins of this family, the $SHMT_{Sth}$ active site can present either an “open” or a “closed” conformation, allowing the reagents and cofactors to move in and out of the active site. In the *apo*-structure the enzyme is found in an "open" conformation, due to the absence of the reaction substrates. In the structures of binary complexes the conformation is “closed, as further explained and illustrated by figures in the next paragraph.

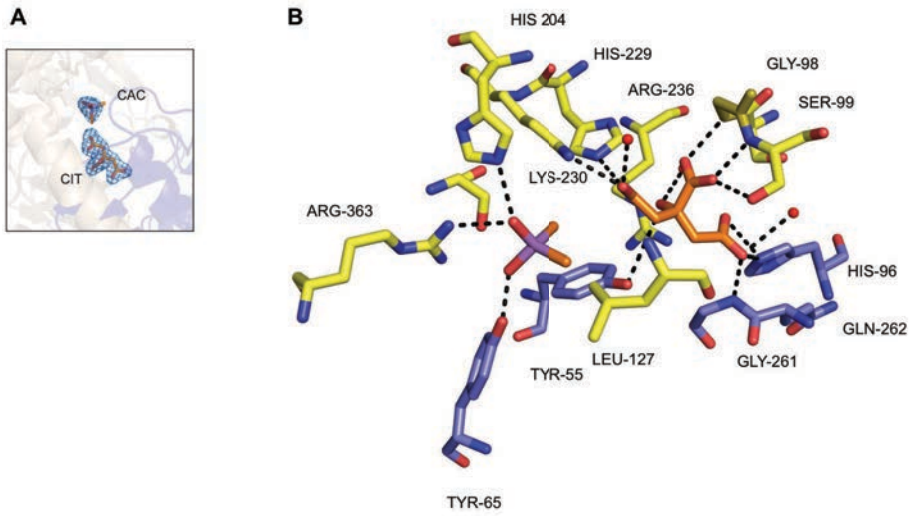


Figure 16: The active site of $SHMT_{Sth}$ *wild-type apo*-form

(**A**) Diagram illustrating the electron density map ($2F_0–F_c$ at 1σ) corresponding to cacodylate (CAC) and citric acid (CIT) present in the active site of the *apo*-form structure.
(**B**) View of the active site architecture depicting key residues coming from the chain A (in olive) and chain D (in blue). Hydrogen bonds are shown as black dotted lines and waters are shown as red spheres.

1.9.2 Overall structure of $SHMT_{Sth}$ *wild-type* binary complexes (WTgly and WTLThr)

In contrast to the *apo*-structure, both binary complex structures contain just two chains (A-C) in the asymmetric unit, each one consisting of 404 residues; the initial 6 residues and His-tag at the N-terminus are not visible in the electron density map and, for this reason, are not included in the final models. The final models contain residues 7-411, two molecules of PLP bounded to glycine and L-threonine respectively, and between 350 and 400 water molecules.

The Ramachandran plot for both is indicative of a good geometry of the backbone torsion angles as it shows that 97% of all residues fall into the core region, and only four non-glycine, non terminal amino acid residues lie outside the core or most restrictive extension region of allowed values. The residues are Lys-230 and Thr-356 in both chains.

The structures of the two binary complexes are virtually identical to that of the *apo*-form structure except for local conformational changes. When superimposed with the *apo*-structure, the binary complexes show an r.m.s.d. of 0.40 and 0.41 Å respectively, over the 410 Cα carbons. The overall structures of the two binary complexes superposed with the *apo*-form are shown in the picture above (**Fig. 17**). The crystal structures of the complexes clearly establish local, yet significant conformational changes with respect to that of the enzyme in its *apo*-form, involving the PLP ring and side-chain atoms in the active site region, which also influence the crystal packing organization.

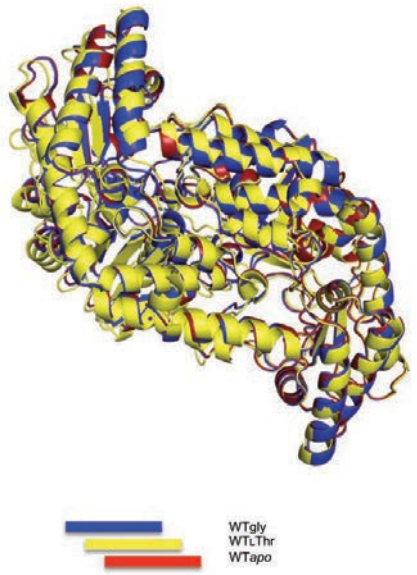


Figure 17: Dimeric quaternary structure of binary complexes superimposed

Cartoon representation of the WTgly (in blue) and WTLthr (in yellow) dimers, superposed to the WT*apo* structure (in red) The structure appear to be very similar.

1.9.2.1 $SHMT_{Sth}$ *wild-type* binary complexes (WTgly and WTLThr): catalytic site architecture and proposed catalytic mechanism

Two important regions of the molecule participate in the formation of the active site: the first region provides key residues for PLP recognition and binding; the second region, located on the opposite side of the active site, belongs to the partner subunit and contributes to the formation of the substrate-binding cleft.

As reported by Vidal *et al.* in 2005 and previously seen in the introduction of this chapter, $SHMT_{Sth}$ shows LTA activity.

In the WTgly structure, PLP was found in the active site as an external aldimine forming a Schiff base with the amino group of glycine substrate. In the presence of L-threonine substrate and PLP, the enzyme catalyses the retro-aldol cleavage reaction. Therefore, in the WTLthr structure, at equilibrium a mixture of two different external aldimine species are visible in both active sites in similar ratios. Refinement of the disordered structure against the experimental diffraction data shows that the external aldimine in which the C4' of PLP is bound to the amino group of the L-threonine residue is present to an occupancy of 40% whereas the external aldimine in which the C4' of PLP is bound to the amino group of the glycine, derived from the catalysed cleavage of L-threonine into PLP-glycine quinonoid complex and acetaldehyde constitutes 60%. In this case the electron density corresponding to the acetaldehyde is not visible in the map, since it is released to the bulk medium and/or too mobile and disordered.

In summary, the binary complex PLP and L-threonine forms an internal aldimine with the catalytic lysine 230; subsequently, the enzyme catalyses the retro-aldol cleavage of L-threonine to glycine.

The carboxylate group of the threonine interacts, through one oxygen (O), with Arg-363 from monomer A (2.7 Å) and the imidazole group of His-204 from monomer C (2.8 Å) and through the other oxygen (OXT) with the phenyl group of Tyr-65 from monomer C (2.6 Å), Ser-35 from monomer A (2.88 Å) and Arg-357 from monomer C (3.07 Å). The carboxylate group of the glycine shares the same interactions. The hydroxyl methyl group of threonine interacts with Glu-57 from monomer C (2.8 Å). All the interactions, made by the amino acid substrates and PLP cofactor with the enzyme backbone, are shown in the **Fig.18**.

The overall environment of the PLP in the binding pocket is very similar to that found in other SHMTs (Trivedi et al., 2002) (Scarsdale et al., 1999). The interactions seem to be conserved during the different steps of catalytic reaction, with the exception of the covalent bond to the Lys-230 and the flip of the backbone of the adjacent His-229 residue.

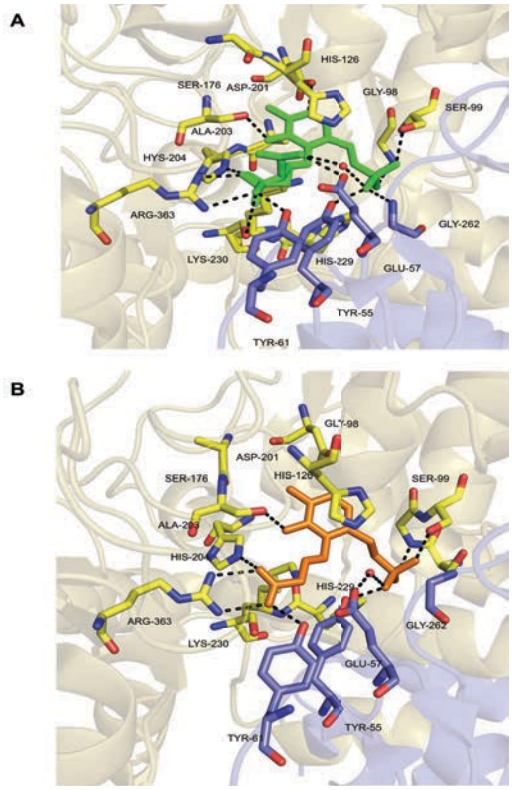


Figure 18: Ligand interactions inside the active site

(**A**) Active site geometry of $SHMT_{Sth}$ in complex with gly depicting the amino acids involved in ligands bounds. On the right side of the picture, ligands plot of the PLP/L-Gly external aldimine interactions.
(**B**) Active site geometry of $SHMT_{Sth}$ in complex with L-Thr and corresponding ligands plot.

Breaking of the internal aldimine results in the reorientation of the catalytic Lys-230 and His-229 side chains, as evidenced in the structure: substrate binding correlates with the peptide flip in residue His-229 and the disorder in Lys-230. The lysine orientation through the PLP Cα correlates with the presence of L-threonine bound to the cofactor. (**Fig.19**)

The correlated carbonyl flip of the backbone residue His-229 is also observed. Moreover the WT*apo* and WTgly carbonyl disposition correlates with the *apo*-structure of *bs*SHMT (PDB code:1KKJ), whereas the flipped carbonyl corresponding to the PLP/Lthr complex in the WTLthr structure corresponds to the disposition observed in the *halo*-structure of bsSHMT (PDB codes: 1KKP, 1KL1, 1KL2).

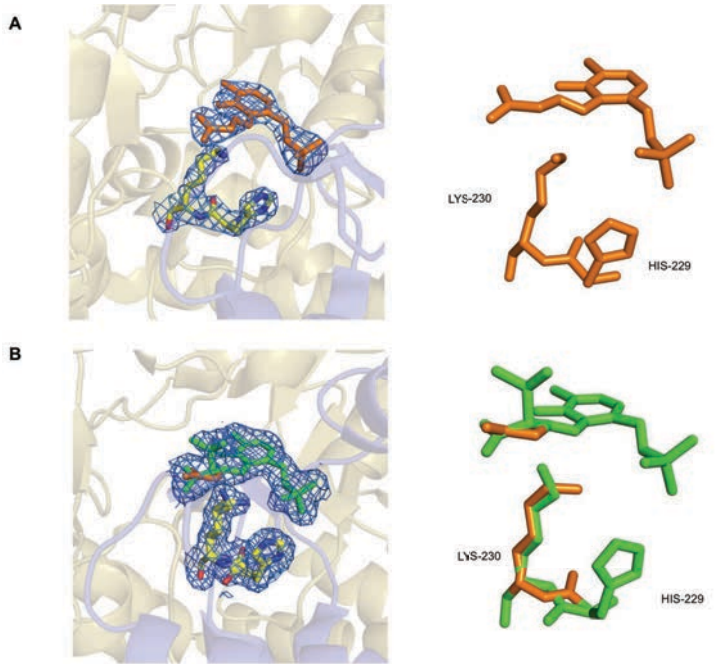


Figure 19: Electron density around the ligands at the active sites of the $SHMT_{Sth}$ binary complexes and peptides flip

Diagram illustrating the electron density map ($2mF_0–DF_c$ at 1σ) around the glycine external aldimine in the WTgly structure **(A)** and PLP/L-Thr in the WTLthr **(B)**.
In the WTLthr structure, both complexes (PLP/Gly and PLP/LThr) are present in the active site at the same time with partial occupancy. The backbone flips in residues K230 and H229 are also shown under both structures, using the same colours of the corresponding reaction intermediates: in orange gly-external aldimine and in green L-Thr-external aldimine.

The structural studies of the complexes indicate concerted changes in several amino acid residues in the catalytic mechanism of the enzyme and the compared studies of *apo*-$SHMT_{Sth}$ and $SHMT_{Sth}$ binary complexes has further helped in probing the role of each residue present in the active site.

As in other fold-type I enzymes, each monomer is composed of a small and large domain, with substrates bound at their interface. The bound PLP-Gly/L-Thr in their complexes make several interactions with both monomers (A, C) in the protein (**Fig.20**), showing in the PLP-Gly complex all corresponding interactions, except for those formed by the absent hydroxyl group of threonine.

In the WTLthr structure the hydroxyl group of threonine is in close proximity to the carboxylate group of Glu-57 from monomer C (2.79 Å) and the imidazole group of His-126 from monomer A (3.24 Å), interacting with them through hydrogen bonds through a crystallographic water, W794A, (2.18 Å), that also interacts with the phosphate group of PLP (2.99 Å).

The retro-aldol mechanism for the LTA reaction of L-Thr requires a suitably positioned residue or species in the active site that could abstract a proton from the hydroxy group of L-Thr, to activate the formation of quinonoid intermediate (**Scheme 3**). It has been suggested in previously studies on SHMTs from other sources that Glu-57 (nomenclature in $SHMT_{Sth}$) at the active site can carry out this function (i.e. the essential acid-base residue) (Rajaram *et al*., 2007). Studies conducted on serine hydroxymethyltransferase from *Escherichia coli* (*e*SHMT) variants show that actually this glutamate (Glu-53 in *e*SHMT), is not involved in this function.

The hydrogen of the hydroxyl group of L-threonine (H-O-CH_2-Cα)-PLP external aldimine complex, which has to be abstracted to form the quinonoid intermediate is located at 2.79 Å from the Glu-57 (4.4 Å in the computational modelling) and at 2.18 Å from a water molecule. This water molecule also establishes H-bond with the phosphate group of PLP.

For the catalytic reaction to occur, it is necessary to place the L-Thr-PLP external aldimine in a spatial conformation that allows the hydrogen of the hydroxyl group of L-threonine-PLP in close proximity to an acid-base residue, which could be Glu-57 or phosphate activated water molecule in WTLthr.

Mutation studies conducted in our research group on residue Glu-57 (E57Q, E57L, E57V, E57K) demonstrated that this residue, Glu-57, is not essential for catalysis to take place. In fact, the substitution of the glutamate in this position by amino acids whose lateral chain cannot provide its acid-base properties (Gln, Leu, Val) does not affect the catalytic reaction.

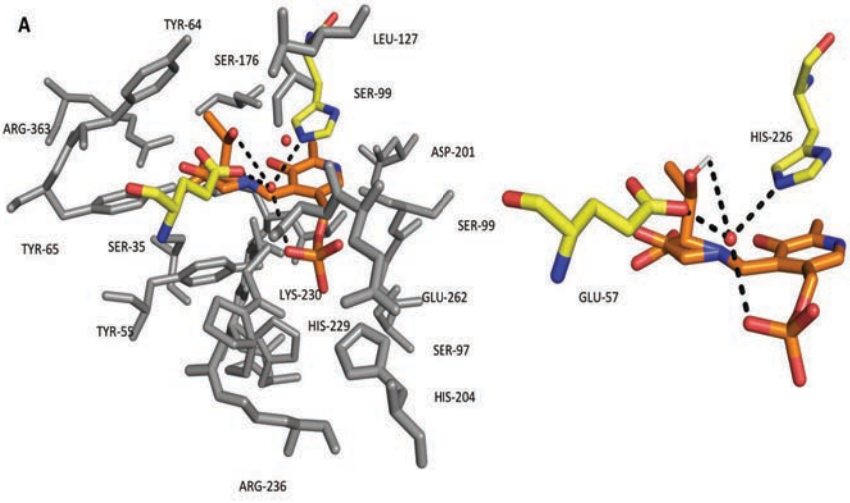
Variant of $SHMT_{Sth}$	Aldol activity (%)
WT	61
E57Q	95
E57L	58
E57V	91
E57K	19

Table 3: Screening of the variants in position 57

Screening of $SHMT_{Sth}$ variants in positions E57 including the *wild-type* as catalyst for the aldol addition of glycine to dimethoxylacetaldehyde.

Considering the results obtained from the mutational studies on residue Glu-57, along with the computational modelling experiments (Hernandez *et al*., 2014) and the crystallographic analysis of the structure, we can conclude that the acid-base residue involved in the

proton abstraction could plausibly be the water molecule that transfers its proton to the phosphate ester group of PLP (step I in **Fig. 20B**) and then it abstracts the proton from the hydroxyl group of threonine in the external aldimine, leading to the formation of the quinoid intermediate (step II **Fig.20B**), as also observed in L-threonine aldolase from *Aeromonas*



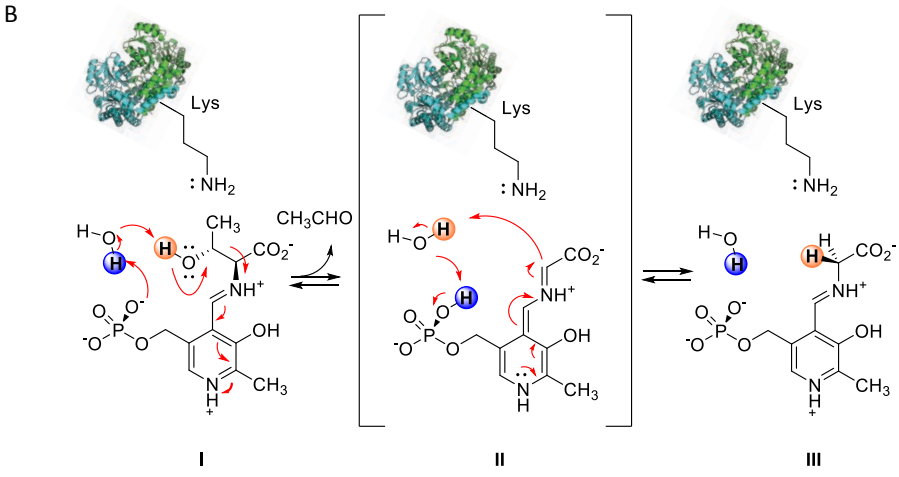


Figure 20: Ligand interactions inside the active site

(**A**) Active site of WTLThr. On the right side of the picture, H-bond interactions (dashed black lines) to the catalytic water (red sphere) the putative acid-base residue and the phosphorylated group of L-Thr-external aldimine, Glu-57 and His-226 residues.
(**B**) A schematic representation of the proton abstraction, mediated by the catalytic water, on the hydroxylmethyl group of L-threonine-external aldimine (step I), which triggers the quinonoid intermediate formation (step II). The water molecule is deprotonated by the phosphorylated group of PLP cofactor (proton shown in blue in step I). The glycine external aldimine is formed (step III).

The swap from the "open" to "closed" conformation, involving the amino acids Thr122 to Thr133, is evident in the structures in complex with substrates: in the *apo*-structure the active site is exposed to give access to the substrates for the physiological reaction to occur (**Fig.21**). Upon binding of the amino acids to the enzyme, the closed form prevents the reagents and the cofactor from moving out from the active site. Formaldehyde is a poison for the cell; so closing the site prevents a potential damaging side reaction. Moreover the closed form blocks the unwanted reaction and enhances the rate of catalysis by a correct alignment of the catalytic residues; while, in the “open” active site conformation, the enzyme loses control of proton placements, resulting in the non-physiological side reaction (Contestabile et al., 2000).

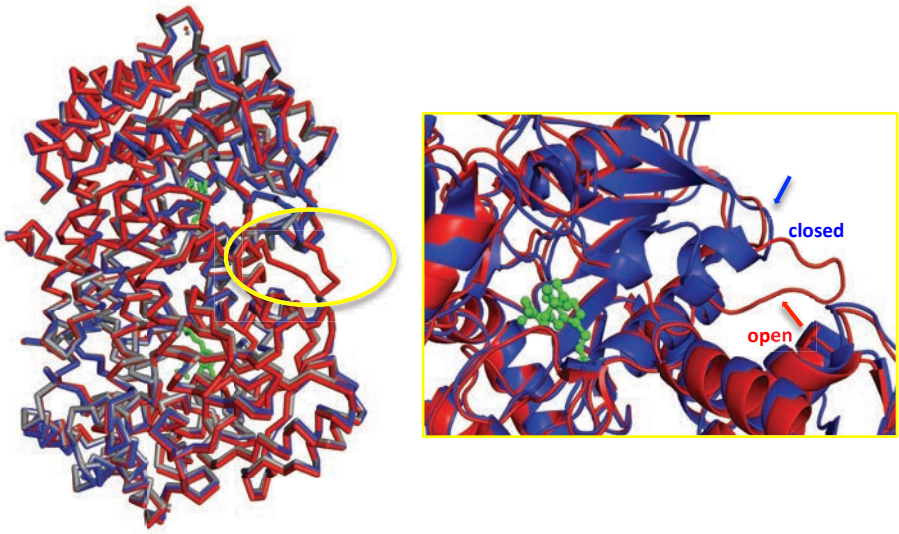


Figure 21: "Open" and "closed" conformation

(**A**) Superposition of binary complexes (WTLthr in gray and WTgly in blue) and the *apo*-structure (in red) showing differences in the conformation of the catalytic task channel. A remarkable conformational difference between the binary complex and the *apo*-structure is indicated by the yellow oval, showing that the "open" conformation is present in the *apo*-form and the "closed" conformation is present in the *halo*-forms of the protein. (**B**) Zoom into the region, showing the different conformation in more detail. As evident from the picture, the alpha helix unfolds into a coil in the "open" conformation (red), whereas it is well-structured in the "closed" conformation (blue).

Superposition of the three structures indicates that the "open" conformation of the dimer is achieved through a concerted rigid-body movement of the connecting loop between two helices present in the large N-terminal domain. The magnitude of the fluctuations for the N-terminal and PLP-binding domains is much higher in *apo*-$SHMT_{Sth}$ than in *holo*-$SHMT_{Sth}$ structures, as also suggested by crystallographic isotropic displacement parameters (B-factors) for the amino acids in

these regions, where which B-factors in the closed conformation structures are in average 28, whereas for the ones in the open conformation the average is 65. Moreover, in the "open" conformation, the PLP sensor helix (residues 290-303) and the surrounding regions are clearly more mobile after the opening transition, which exposes this region to the solvent, whereas in the closed dimer, this helix is buried and stabilized by the presence of PLP.

1.10 Structures of the $SHMT_{Sth}$ Y55S and Y55T variants

As explained in the Introduction section of this thesis work, particular mutations at position 55 (Y55T, Y55C, Y55S) of the $SHMT_{Sth}$ lead to enhanced catalytic activity toward a wider range of nucleophile substrates, such as D-Ala and D-Ser with respect to the *wild-type* enzyme.

Crystallization studies on these $SHMT_{Sth}$ variants were carried out to determine the amino acid binding pocket and elucidate the reasons of such enhanced activity. For the crystallographic studies, the mutant Y55S was selected first, considering its lower reaction rate compared with the Y55T variant. In fact, a lower reaction rate is of practical convenience in the case of crystallization studies, as it facilitates to isolate intermediate reaction steps. Thus, the Y55T variant was crystallized in the presence of D-serine. Other substrates selected for the structural studies were: D-threonine and L-threonine.

In total we have obtained and studied five structures:

- Y55S mutant in complex with L-threonine (L-threonine external aldimine, **Y55SLThr**)

- Y55S mutant in complex with D-serine (D-serine internal aldimine, **Y55SDSer-PLP**)
- Y55S mutant in complex with D-serine (D-serine external aldimine, **Y55SPLS**)
- Y55S mutant in complex with D-threonine (**Y55SDThr**)
- Y55T mutant in complex with D-serine (internal and external aldimine, **Y55TPLS/DSer-PLP**)

1.10.1 Y55S SHMT$_{Sth}$ L-threonine external aldimine (Y55SLThr)

The overall structure of Y55SLthr is very similar to that of the native structure of the SHMTSth L-threonine external aldimine, existing as a homodimer displaying two-fold symmetry, constituted by two chains (A-C) of 404 amino acids with two catalytic sites. The r.m.s.d. of the corresponding Cα atoms upon superposition of the *wild-type* and mutant structure is 0.261. As in the case of *wild-type* structures, neither the initial 6 residues from the N-terminal nor the N-terminal His-tag are visible in the electron density map and thus are not included in the final models. The final model contains residues 7-411, two molecules of PLP bounded to L-threonine (in its external aldimine) and 392 water molecules.

The Ramachandran plot shows that 97% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt and Jones, 1996), fourteen non-glycine, non terminal amino acid residues lie within the core or most restrictive extension region of allowed values, and finally six residues are outliers: Lys-230, Asn-316 and Thr-359 in both chains. Lys-230 is very flexible, being involved in the catalytic reaction; Thr-359 is located in close proximity to the active site cleft

that is a very mobile region of the protein; Asp-316 is part of a closed turned. The detected outliers show clear electron density around them.

In the case of this structure, an unambiguous electron density corresponding to the external aldimine (i.e., PLP bound to L-threonine) is evident in the active site, whereas, as explained previously, in the *wild-type* structure a mixture of internal aldimine (PLP-L-Thr) and external aldimine (PLP-Gly) is visible, suggesting that L-threonine was converted into the product glycine under the crystallization conditions. This can be explained by the fact that variants in position 55 do not show retro-aldol activity against L-Thr, as demonstrated by studies conducted in our laboratory, and therefore the conversion of L-Thr to glycine did not occur. The electron density around the threonine for the variant structure is shown in **Fig.22**, compared with the electron density in the corresponding *wild-type* structure. The Lys-230 and His-229 residues showed the same flip of the carbonyl backbone as the one observed in the corresponding *wild-type* structure, which correlates with the presence of L-threonine. (**Fig.22**).

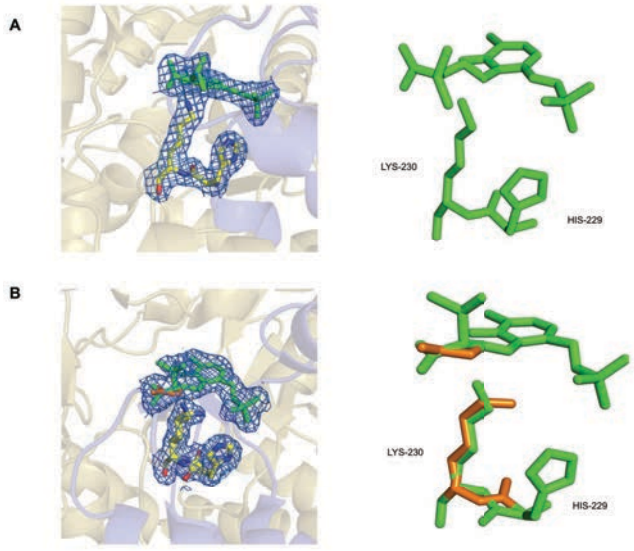


Figure 22: Electron density around the cofactor PLP and substrates at the active sites of the $SHMT_{Sth}$ WTLthr and Y55SLthr with a focus on K230 and H229 flips

(**A**) The electron density map ($2F_0–F_c$ at 1σ) corresponding to PLP/L-threonine substrate in the WTgly structure and PLP/L-threonine in the Y55SLthr. In the case of Y55SLthr the electron density allows placing unambiguously the external aldimine complex, formed by PLP and L-thr and it is also unambiguous around the residues K230 and H229 (**B**), which are present in double conformation in the corresponding *wild-type* structure. As shown, in Y55SLthr the backbone peptide flips in residues K230 and H229 are coinciding with the same that the ones correlated to the presence of L-threonine in the *wild-type* structure.

The mutation does not induce any gross conformational changes in the protein molecule. However, a subtle, yet significant, rearrangement of the catalytic site is evident. The mutation introduced in position 55 entails a loss of the close contact between phosphate group of PLP and the Tyr-55 residue (2.5 Å) seen in the wild type. The hydroxyl oxygen in this tyrosine, not only contacted the cofactor but also the side chains

of His-229 (2.8 Å) and Arg-236 (4 Å). Nevertheless, in the Y55S variant, a structural water is present in the position occupied by this hydroxyl oxygen of tyrosine in the *wild-type* structure, undergoing the same hydrogen bonds to the PLP cofactor (2.5 Å) and His-229 (2.9 Å). The location of this water is further stabilized in the absence of the tyrosine side chain by Glu-57 (3.3 Å), which changes the conformation of its side chain to support the water, while still contacting with the catalytic water molecule (2.9 Å) and the hydroxymethyl group of L-Thr in the external aldimine complex (3.2 Å). Glu-57 is moved towards the serine in position 55, allowing the PLP to be held in place. (**Fig.23**).

The electronic and steric changes observed in the active site can explain the lack of retroaldol activity toward L-Thr and in fact we have not been able to isolate an intermediate with reaction products, unlike what happened in the *wild type* structures.

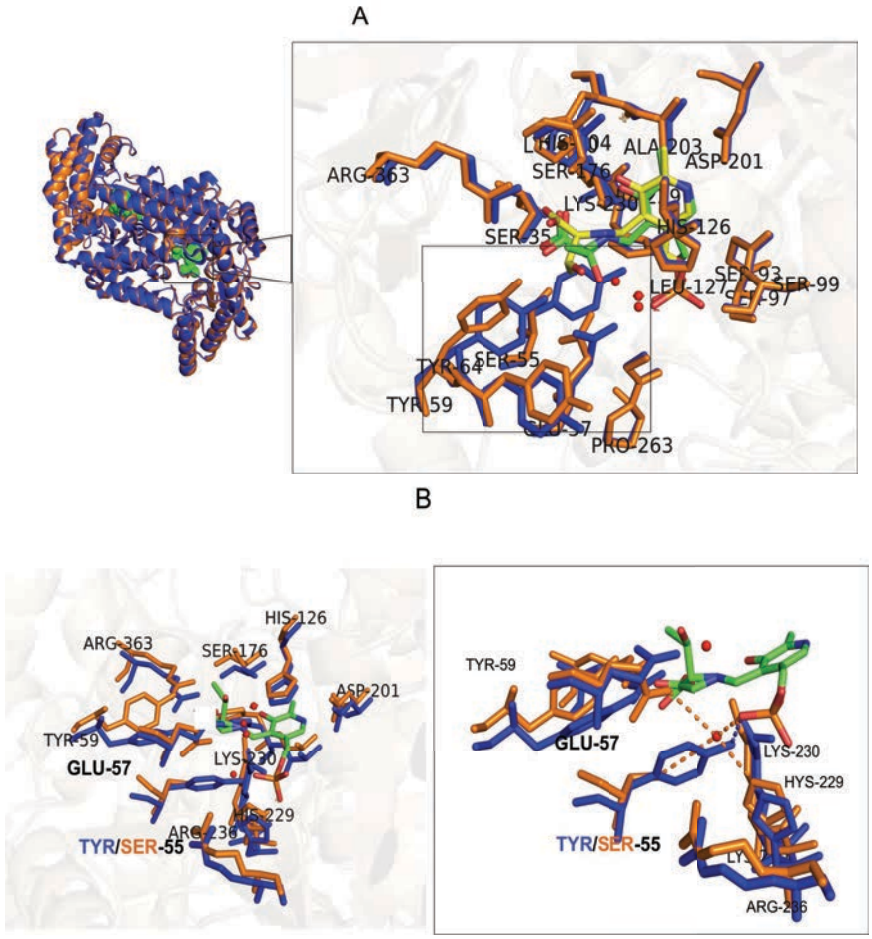


Figure 23: Comparison between *wild-type* structure and Y55S mutant in complex with L-Thr

(**A**) Superposition of the two structures, shown in blue (WTLthr) and in orange (Y55SLthr), evidences their similarity. On the right side, the overlay of the active site regions of the WTLthr (in blue) and Y55SLthr (in orange) reveals some local conformational changes, directly connected to the mutation introduced, as further commented in the paragraph. The ligands are shown in yellow and green respectively. The mutation generates a space in the active site, due to the absence of hydroxymethyl group of tyrosine (substituted by serine) and the re-allocation of Tyr-64, Tyr-59 and Glu-57 (evidenciate in the black square). (**B**) Another perspective, showing the water occupying the position of the hydroxylgroup of the tyrosine in Y55SLthr and a zoom on the amino acids involved in the H-bonds with this molecule.

The bound PLP cofactor shows exactly the same interactions illustrated for the *wild-type* structure with backbone residues. The orientation of the PLP in Y55SLthr is equivalent to that of the *wild-type* binary complex, except for minor movement, depending on the different stages at which the two structures are in the reaction mechanism. The most consistent differences are seen in the active site and around the mutated residues in position 55. Substitution of the tyrosine in position 55 by serine generates a space in the active site that is occupied by the hydroxyl group of the L-threonine substrate, which is now more distant from the catalytic water molecule (i.e., W45E in this structure) and from the Gly-57 (**Fig.24**). This may explain the absence of retroaldol activity of Y55S variant against L-threonine: the cleavage of the external aldimine to quinonoid complex is not possible anymore and the reaction halts in this step (**Fig.24**). Another interesting change that can explain the lack of retroaldol activity for the variants is the orientation of the L-threonine in the external aldimine, which is perpendicular to the PLP plane in the case of the *wild-type* structure and parallel to it in the case of Y55S variant (**Fig.26**). Dunathan postulated in 1966 (Dunathan, 1966) that the topology of the amino acid aldimine determines the bond to Cα that would be broken. He suggested that that bond must be situated so that it will align perpendicularly with the pyridine ring of the cofactor in the transition state of the reaction. The ensuing carbanion is stabilized by conjugation with the extended pi system. This hypothesis was later confirmed when the structure of the aspartate amino-transferase/phosphopyridoxal aspartate complex was solved. The orientation of the L-threonine, which now occupies the space generated by the mutation, impedes the reaction to occur.

No major changes are evident elsewhere in the structure.

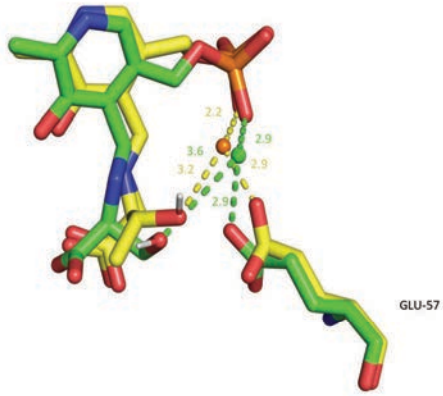


Figure 24: WTLthr and Y55SLthr catalytic *triad* comparison

Bond distances between E57, catalytic water, PLP phosphate group and hydroxyl group of L-Thr in Y55SLthr structure are shown in this picture (in green) compared with the bond distances between the same species in the *wild-type* structure (in yellow).

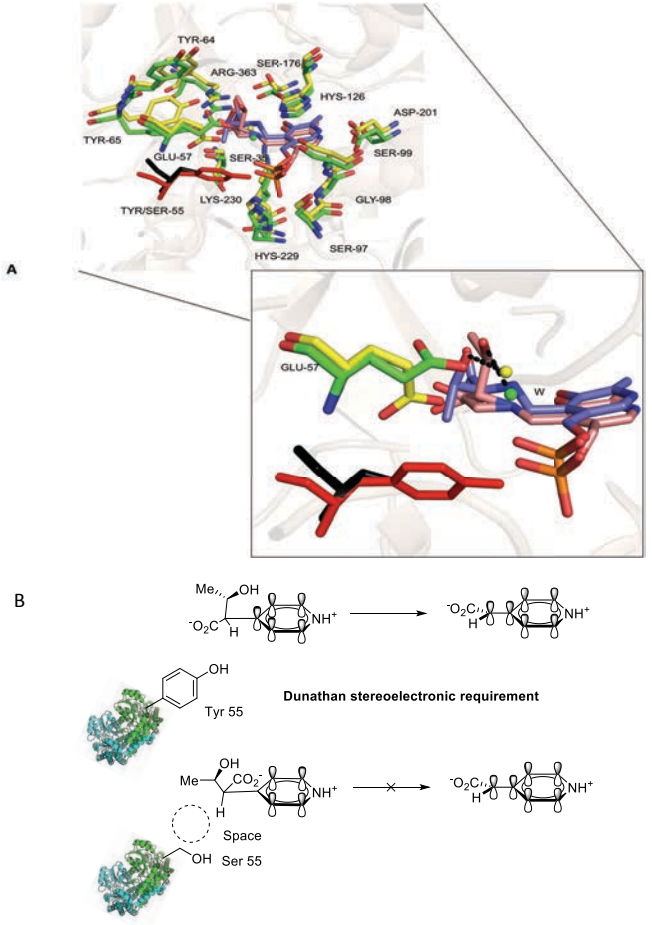


Figure 25: Dunathan's stereoelectronic hypothesis

(A) Superposition of the L-Thr external aldimine in WTLThr and Y55SLThr structures shows the different orientation of the amino acid substrates in the two structures with respect to the PLP plane. On the right side of the picture, a zoom of the reagion, depicting the orientation of the binary complexes in the two structures and the the coordination of the hydroxyl group of L-Thr with phophorilated group of PLP and Glu-57. The mutation in position 55 is also shown. (**C**) The Dunathan stereoelectronic hypothesis. Substrates (L-threonine in the figure) are bond to PLP such that the bound to Cα that is to be broken is aligned with the pi orbitals of the cofactor in WTLThr structure. The substrate orientation is completely modified by the mutation in position 55, which generates a space that will be occupied by the substrate. The Cα is not aligned and the reaction is stuck.

1.10.2 Y55S $SHMT_{Sth}$ D-threonine complex (Y55SDThr)

D-threonine was also screened as putative nucleophile substrate for the library of mutants (94 clones) of $SHMT_{Sth}$ constructed in our laboratory (Hernandez et al., 2015b), but no active clones were found for this substrate. In order to explain why D-threonine was not a substrate for $SHMT_{Sth}$, co-crystallization of Y55S mutant with D-threonine and PLP was designed.

The overall structure of Y55SDthr is very similar to that of the other structures described with the asymmetric unit containing two dimers, constituted by four chains (A-B-C-D) of 404 amino acids with four catalytic sites. The r.m.s.d. of the corresponding Cα atoms upon superposition of the *wild-type* and mutated structure is 0.261. As in the case of all the structures described in this thesis work, neither the initial 6 residues from the N-terminal nor the attached His-tag are visible in the electron density map and then are not included in the final models. The final model contains residues 7-411, four molecules of D-threonine and 78 water molecules.

The Ramachandran plot shows that 97% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt & Jones, 1996), fourteen non-glycine, non terminal amino acid residues lie outside the core or most restrictive extension region of allowed values, and finally six residues are outliers: Lys-230, Asn-316 and Thr-359 in both chains.

The electron density for the ligand is very clear, allowing the substrates to be placed unambiguously in all the chains (**Fig.26**)

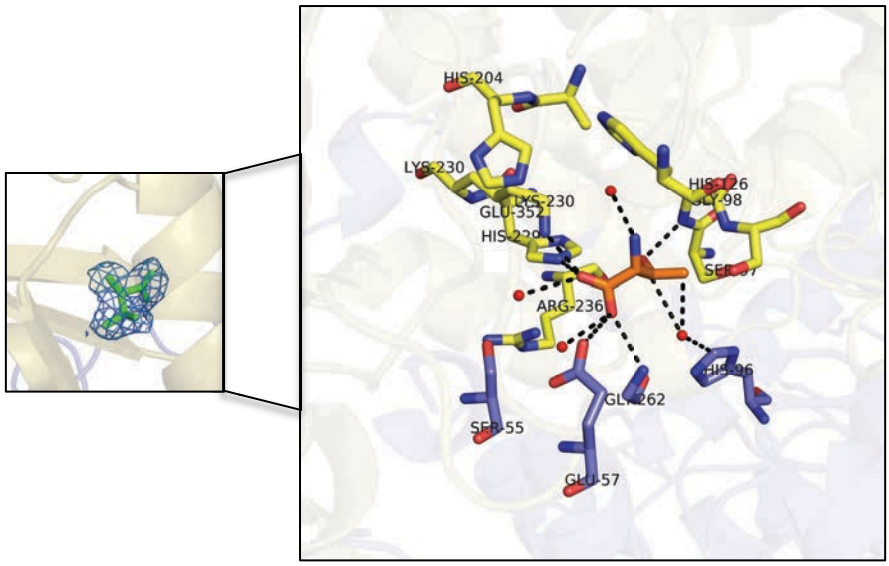


Figure 26: Electron density around the D-Thr at the active sites of the $SHMT_{Sth}$ WTDthr and its interactions

The electron density map ($2F_0–F_c$ at 1σ) corresponding to D-threonine allows placing unambiguously the substrate in all the active sites. On the right site of the picture, a zoom in of the catalytic site is shown.

The D-Thr occupies the position generally occupied by the phosphorylated group of PLP. The hydroxyl group is bounded to Gly-98 (2.8 A), His-229 (3.1 A) and to a water molecule (3.3).

The carboxylate group of the threonine interacts, through one oxygen (O), with Lys-230 from monomer A (2.6 Å) and the imidazole group of His-229 from monomer A (2.1 Å) and through the other oxygen (OXT) with the Gly-262 from monomer C (3 Å). The hydroxyl methyl group of threonine interacts with Gly-98 from monomer A (2.9 Å) and

His-229 from monomer A (3.4 Å). All the interactions made by the amino acid substrates D-Thr and enzyme backbone are shown in the **Fig.26**.

The threonine was found where the phosphate group of PLP is located in the other structures where the cofactor present (**Fig. 27**).

Despite the absence of PLP cofactor, the catalytic lid is in its "closed" conformation. Considering that PLP binding induces the "open"/"closed" transition (Giardina et al., 2015), we suppose that the D-threonine is miming the PLP cofactor after displacing it out of the active site. Closure of the active site would obstruct reconstitution of the active enzyme, even if PLP was available from the reaction medium.

Due to the absence of cofactor in the active site, the enzyme is inactive with this substrate.

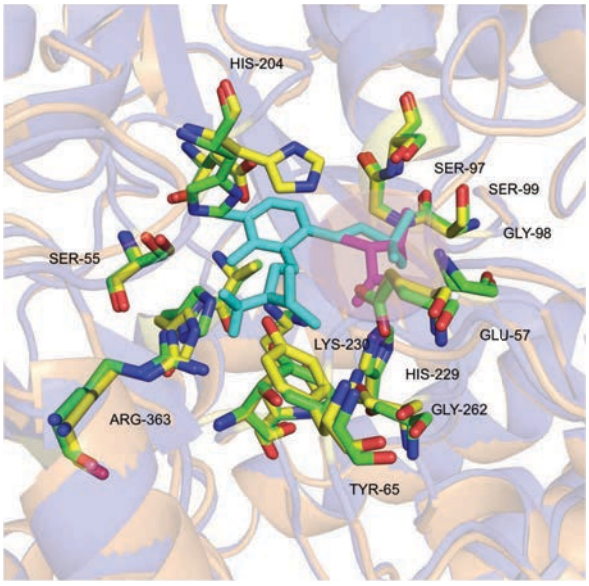


Figure 27: Y55SDThr and Y55SLthr superposition

The D-threonine is found at the position occupied by the phosphorylated group of PLP (red circle).

1.10.3 Y55S $SHMT_{Sth}$ in complex with D-Serine

D-serine internal aldimine (Y55SPLS) and D-serine external aldimine (Y55SDSerPLP)

It has been shown that Y55S $SHMT_{Sth}$ was an efficient catalyst for aldol additions of D-serine to aldehydes, while the *wild-type* enzyme is inactive with this substrate (Hernandez *et al*., 2015).

In order to elucidate the basis of this catalytic activity, crystals of the mutant Y55S were obtained in the presence of PLP and D-serine and two consecutive reaction steps have been isolated, analysing the corresponding structures: the internal aldimine with PLP covalently bound to the catalytic Lys-230 and D-Ser also present in the active site (electron density corresponding to D-Ser is also present in the active sites of A and B chains) and the following step with external aldimine, consisting of PLP bounded to D-serine (1 to 2 and 2 to 3 species of **Scheme 2** in the introduction of this chapter). As mentioned before, the two structures crystallize in different space groups and have different asymmetric unit content: Y55SPLS crystallyses in $P4_12_12$ with two copies (chains A-C) in the a.u. and Y55SDserPLP in $C2$ space group with four copies in the a.u. (A-B-C-D). Remarkably, the open/closed conformational change of the dimer does not underlie the different crystal packing: both structures are in closed conformation.

The r.m.s.d. of the corresponding Cα atoms upon superposition of the *wild-type* and mutant structure is 0.25. As in the case of the structures described up to now, neither the initial 6 residues from the N-terminal nor the N-terminal His-tag are visible in the electron density map and are absent from the final models. The overall structures are shown above above superposed to each other. (**Fig 28**).

The Ramachandran plot for both shows that 97% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt & Jones, 1996), and only six non-glycine, non-terminal amino acid residues lie outside the core or most restrictive extension region of allowed values.

In both structures the catalytic pocket is in its "closed" conformation, as just mentioned, and that is of great interest in order to confirm the hypothesis that the conformational change happens when the substrates are both in the active site and not just on the verge of catalytic reaction. (**Fig. 28**).

The environment of the cofactor and substrate in the two structures is similar with some interesting exceptions. The interactions seem to be conserved during the different steps of catalytic reaction, except for the ones involving the serine, which occupies different positions during the two steps, internal and external aldimine.

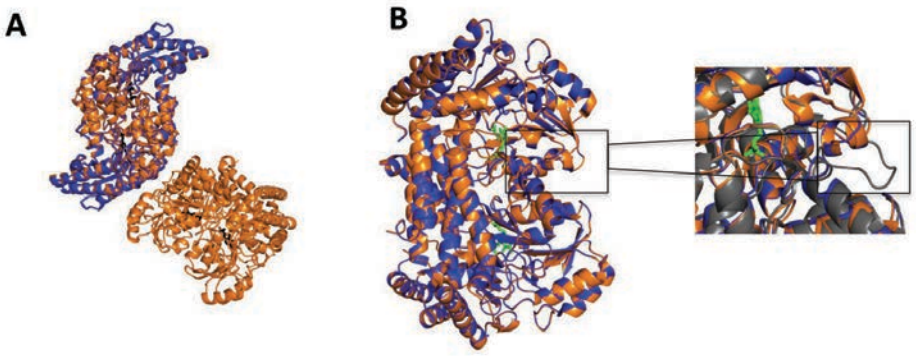


Figure 28: Overall view of superimposed Y55S structures in complex with D-ser and their catalytic task conformation

(**A**)Structural superposition of Y55SPLS (orange) and Y55SDserPLP (blue), showing a very similar overall conformation. (**B**) Overlay of the obligate dimer of Y55S D-Ser complexes showing the catalytic task conformation: a "closed" conformation is evident (black rectangle) in both complexes. On the right side a comparison with WT*apo* structure (in grey), which is in "open" conformation, is shown, in order to highlight the difference between both conformations.

The X-ray structure explains why D-serine is not a substrate for SHMT wt and, on the contrary, it is a good nucleophile substrate for SHMT Y55S or Y55T variants. As evident from the structure, the α-hydrogen of D-Ser in Y55S and Y55T variants is accessible to the catalytic water that, in conjunction with the phosphate ester of PLP, can abstract the proton and triggers the formation of the nucleophilic quinonoid intermediate. In the *wild-type* structure, in agreement with computational models (Hernandez *et al*., 2014), the hydroxyl group of D-Ser would adopt an opposed orientation, stabilized by H bonds with residues Glu-57 and Tyr65 and hindering the access to the H-Cα atom and the reaction is stalled.

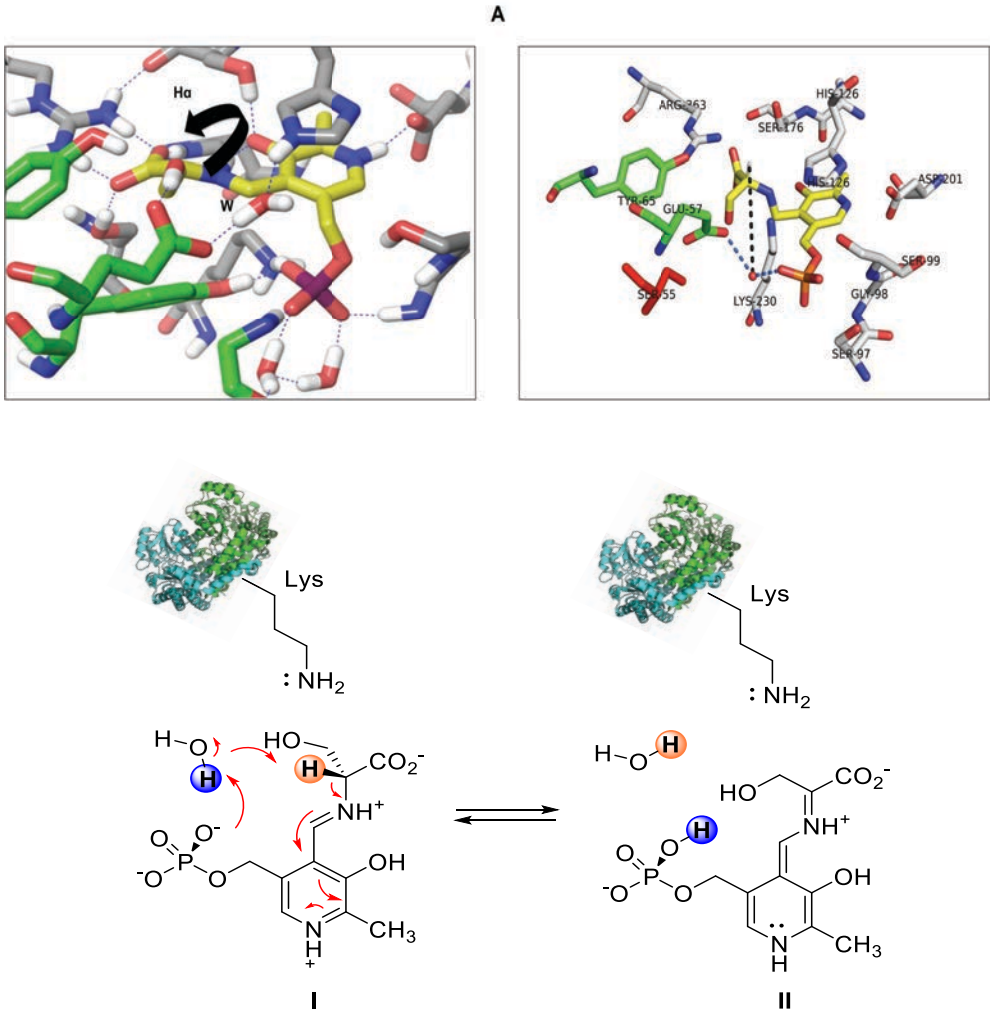


Figure 29: Computational model SHMT$_{Sth}$ D-Ser external aldimine VS X-ray crystallographic structure of Y55S D-Ser external aldimine

(**A**) Computational model of the active site of SHMT$_{Sth}$ *wild type* with D-Ser external aldimine. In this model as evident that the α-hydrogen is inaccessible to the catalytic water. (**B**) X-ray structure view of the SHMT$_{Sth}$ Y55S variant active site with D-Ser-external aldimine in it. The mutation introduced generates a space occupied by the hydroxymethyl group of the amino acid substrate and its α-hydrogen is now not accessible to the catalytic water. (**C**) Schematic view of the aldol reaction catalysed by D-serine helps to understand all the proton transfer during the two reaction steps.

1.10.4 Y55T SHMT$_{Sth}$ in complex with D-Serine (Y55TPLS/DSer-PLP)

Crystals of the mutant Y55T were obtained in the presence of PLP and D-serine and, interestingly, two consecutive reaction steps have been described, taking place at the same time in the same crystal, thanks to the co-existence of the two corresponding reaction intermediates in different active sites: the internal aldimine with PLP covalently bound to the catalytic lysine 230 and D-serine also present in the active site, and the following step with external aldimine, consisting of PLP bounded to D-serine. (**Fig.30**)

The electron density for the substrates was very clear, allowing the ligands to be placed unambiguously.

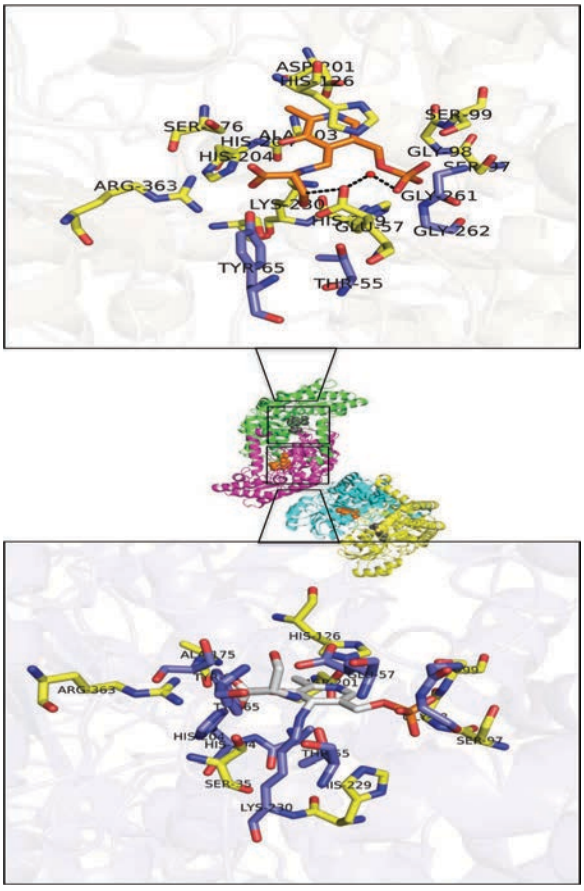


Figure 30: $SHMT_{Sth}$ Y55TDser variant active sites

(**A**) Representation of the Y55TDser structure active site with external aldimine complex. The complex is shown in orange (external aldimine) and gray (internal aldimine), the amino acids in the active site represented as sticks in yellow from chain A and blue from chain C of the molecules. For the sake of clarity, only residues contributing to the ligands bound are shown and further explanation is present in the paragraph. On the right side of the picture, scheme depicting the H bonds (dash lines) observed in the active site. The external aldimine is displayed with bold lines, key water molecule as small sphere.
(**B**) Representation of the Y55TDser structure active site with internal aldimine complex and D-Ser substrate and corresponding H bonds scheme The representation code is the same used in (A).

The overall structure of Y55TDser is very similar to that of the other structures described with the asymmetric unit containing two dimers, constituted by four chains (A-B-C-D) of 404 amino acids with four catalytic sites. The rmsd of the corresponding Cα atoms upon superposition of the *wild-type* and mutated structure is 0.19. As is common to all the structures described in this chapter, neither the initial 6 residues from the N-terminal nor the N-terminal His-tag are visible in the electron density map and then are not included in the final models. The final model contains residues 7-411, two molecules of PLP bounded to Lys-230, in its internal aldimine form, two molecules of D-Ser and two molecules of PLP/D-Ser complex, in the external aldimine, form and 78 water molecules.

The Ramachandran plot for shows that 97% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt & Jones, 1996), fourteen non-glycine, non terminal amino acid residues lie within the core or most restrictive extension region of allowed values, and finally six residues are outliers: K230, N316 and T359 in both chains.

The external aldimine complex shows exactly the same interactions, compared with the ones shown by the Y55SDser structure, described in the previous paragraph. A very interesting exception consists in the orientation of the side chain of Glu-57 that is, in this structure, to a position closer to the crystallographic water responsible for the proton abstraction (2.8 Å compared to the 3.2 Å in the Y55SPLS). This fact can explain the higher synthetic efficiency of this variant compared with the Y55S variant.

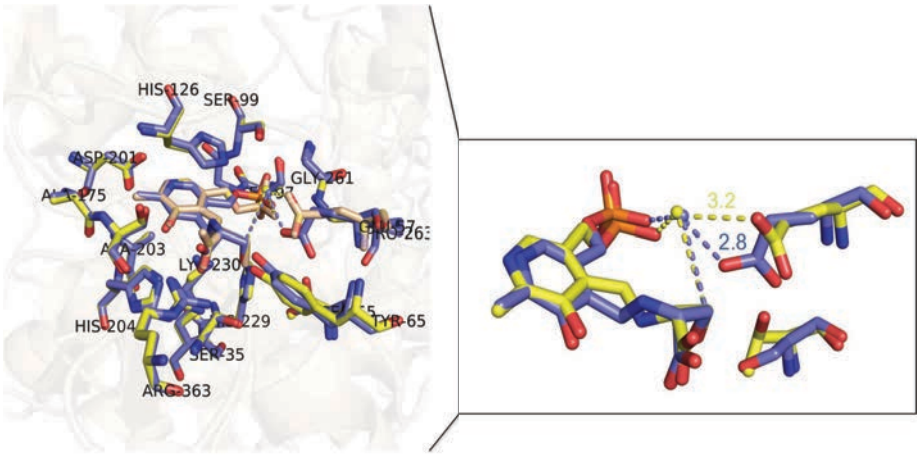


Figure 31: Comparison of the Y55TPLS/DserPLP and Y55SPLS active sites

A comparison between the two variants reveals that, due to the presence of serine or threonine in position 55, the distances between the hydroxymethyl group of serine to Glu-57 and the catalytic water change. In the case of Y55T variant is more closer to them.

The serine substrate in the internal aldimine active sites makes several interactions with both monomers of the active site and, also in this case, they are the same shown by the serine substrate present in the Y55SPLPDser structure and they will not be further commented. No more significant differences can be observed.

In the presence of D-Ser, the PLP undergoes a rotation of approximately 25° to form the external aldimine (around C5-C5' bond of PLP) compared with its internal aldimine form. (**Fig.32**).

A similar rotation of the PLP ring has been reported for *Bacillus stearothermophilus* external and internal aldimine structures (~25°) (Jager et al., 1994).

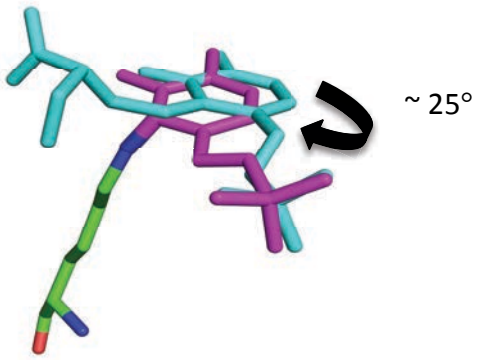


Figure 32: PLP ring rotation

An overlay between the internal aldimine (in magenta) bound to catalytic lysine 230 and the external aldime form (in cyan) complexes shows the change in orientation of PLP ring.

1. 11 General considerations

Which residues or species are involved in the proton transfer step of catalysis?

In addition to the essential lysine residue, involved in covalent bond with the PLP, other residues like arginine, histidine, glutamic acid and tryptophan are important for catalysis and elucidating their role is of a great interest for protein engineering studies.

The structure of an increasing number of SHMTs and their corresponding variants from different organisms, available in PDB, have helped to precisely interpret many of the mechanistic studies for representatives of this class of proteins. Lacking in these studies is,

however, the identification of a putative base that is required to remove the proton from the serine 3-hydroxyl group in the direction of serine cleavage, and to remove the *pro-S* H-Cα proton of the glycine in the direction of serine synthesis.

In other PLP-dependent enzymes, the catalytic lysine (Lys-230 in $SHMT_{STh}$) is involved in the proton transfer step (Schirch et al., 1993). In fact, in these proteins the proton is removed from the α-carbon of the substrate and the active-site base is the ε-amino group of the Lys residue that forms the internal aldimine with PLP and removes the *R* proton on the α-carbon from the *si* face of C4' (the carbon with the aldehyde group of PLP) (John, 1995). This is clearly not the case with SHMT, since the expelled amino group of the active site Lys is on the *si* face of the α-carbon of the bound substrate and would remove the pro-*R* proton of glycine rather than the pro-*S* proton (the pro-*S* would be removed faster by another species).

In *E. coli* SHMT, Lys-229 (Schirch et al., 1993) does not participate in the proton abstraction step and the same is in the case of SHMT from *Bacillus stearothermophilus*, as demonstrated by the structural studies on mutants in this position (Bhavani et al., 2005). The structures of human and rabbit SHMTs suggest that either the tyrosine residue corresponding to Tyr-65 in $SHMT_{Sth}$, or glutamate residue, corresponding to Glu-57 in $SHMT_{Sth}$, is the putative base. Studies conducted on *e*SHMT Tyr-65 and Glu-57 mutants demonstrate that neither Tyr-65 nor Glu-57 are actually the base involved in the proton transfer. In fact, the complete isomorphism of the mutants with the *wild-type e*SHMT precludes effect from changes in enzyme catalytic properties by modifying these residues (Contestabile et al., 2000). In *bs*SHMT, studies conducted on mutants, also reveal that these amino

acids are not involved in the proton abstraction step (Rajaram et al., 2007) (Pai et al., 2009).

Here we propose that the base responsible for the proton abstraction is a crystallographic water molecule, which appear to be activated by H bonds to the phosphate group of PLP.

The D-serine external aldimine model shows that the empty space generated by the mutation in position 55 is partially occupied by the hydroxymethyl group of D-serine, which is now accessible to the catalytic water for the proton abstraction. In the *wild-type* structure, the D-Ser substrate adopts an opposed orientation with respect to the orientation adopted in the variants structures, stabilized by H bonds with residues Glu-57 and Tyr-65 and hindering the access to the H-Cα atom, which would be at more than 5 Å from the Glu-57 and the catalytic water molecule (Hernandez *et al.*, 2014). The lack of D-Ser reactivity with wild-type $SHMT_{Sth}$ could thus be due to the steric hamper caused by the hydroxyl group of D-Ser, placed between the proposed catalytic water and the H-Cα atom.

Insight into the X-ray structures of the variants allows us to speculate on the reason why the Y55T variant is more active than the Y55S variant. In the Y55T variant, the H-Cα is even closer to residue Glu-57 (2.9 and 3 Å in the Y55SPLS and Y55TDser structures respectively) and to the water molecule (2.9 and 3.5 Å), resulting in a more efficient proton transfer and enhanced catalytic reaction.

The studies conducted previously in our research group have shown that mutation in position 55 affects the LTA activity of the enzyme. X-ray structural analysis sheds lights into the mechanism of the $SHMT_{Sth}$ offering a concrete explanation to the theory postulated by computational modelling analysis. Examination of the active site of

$SHMT_{Sth}$ Y55S and Y55T variants shows that the space generated by the substitution of the tyrosine residue determines a reorientation of the hydroxyl group of the L-threonine substrate in the external aldimine complex, which will be now more distant from the Glu-57 (2.9 and 3 Å) and from the catalytic water molecule. This may explain why the absence of retroaldol activity of Y55X variants was detected against L-threonine: the cleavage of the external aldimine to quinonoid complex is not possible anymore and the reaction stacks in this step, contrarily to that observed with SHMT *wild type.*

Finally, D-Thr is not a substrate for SHMTSth variants because it displaces the PLP cofactor, as demonstrated by the corresponding structure.

Experimental procedures

1.12 $SHMT_{Sth}$ *wild-type* and variants production

The plasmid pQE40-SHMT containing the gene for expression of His-Tagged serine hydroxymethyl transferase from *Streptococcus thermophilus* ($SHMT_{Sth}$) was a generous gift from Departament d'Enginyeria Química of the Universitat Autònoma de Barcelona.

SHMTSth gene mutations were introduced with the QuikChange site-directed mutagenesis kit (Stratagene), using the oligonucleotides listed in Table 3. PCR reactions (50 μL) contained the plasmid pQE40-SHMTSth *wild-type* as template (50 ng), the primers (2 μM), dNTP (250 μM each), Tris-HCl (200 mM, pH 8.8 at 25°C), (NH4)2SO4 (100 mM), KCl (100 mM), BSA (1 mg mL–1), Triton X-100 (1% v/v), MgSO4 (20 mM) and Pfu polymerase (2.5 units). The plasmid was amplified in 24 cycles of 95 ºC for 30 s, 55 ºC for 1 minute and 68 ºC for 12 minutes and finally 68 ºC for 30 minutes. Amplification was checked by electrophoresis in a 1% agarose gel in TAE buffer. The DNA was digested with DpnI (0.5 U) at 37ºC for 4 h to eliminate methylated ancestral template and used directly for transformation into *E.coli* NovaBlue (Novagen) chemocompetent cells. Plasmids were purified with the High Pure Plasmid Isolating Kit (Roche) for DNA sequencing.

Oligonucleotide	Oligonucleotide sequences (5`→3`)
SHMT Y55T 53	GACTAACAAAACCGCAGAAGGCTATCC
SHMT Y55T 35	GGATAGCCTTCTGCGGTTTTGTTAGTC
SHMT Y55X 53	GACTAACAAANNKGCAGAAGGC
SHMT Y55X 35	CTGCMNNTTTGTTAGTCAAGAGC

Table 4: $SHMT_{Sth}$ Y55S and Y55T variants oligonucleotides

Oligonucleotides used in this study for the production of Y55T and Y55S variants are indicated and their mutagenized codons are highlighted in red.

1.13 $SHMT_{Sth}$ protein expression

Cultivation conditions for expression of $SHMT_{Sth}$ and variants were as follows: Plasmids were transformed into *E. coli* strain M-15 [pREP-4] (QIAGEN). 2 µL (usually at 100 ng/µL) of the plasmid were deposited at the bottom of the microtube, onto which 50 µL BL21-DE3 pLys-S strain were added and incubated 30 min on ice. It followed a heat-shock step (1 min at 42º C) and 5 min of incubation on ice. Subsequently, 900 µL of LB medium were added to the transformation reaction and the cells were shacked at 200 rpm for 1h at 37º C. 100 µL of reaction were plated on an agar plate with kanamycin and incubated overnight (O/N) at 37º C. One colony of the transformation plate was inoculated into a pre-culture of 4 ml of LB and shaked at 250 rpm O/N at 37º C. A single colony was inoculated into a 50 ml of LB with ampicillin (100 µg mL–1) plus kanamycin (25 µg mL–1) at 37°C on a rotary shaker at 250 rpm overnight. A final OD600 of 3–4 was usually achieved. An aliquot of pre-culture (12 mL) was inoculated to a 2 L shake-flask with LB medium (400 mL) containing ampicillin (100 µg

mL–1) and kanamycin (25 μg mL–1) and incubated at 37°C on a rotary shaker at 250 rpm. Protein expression was induced with IPTG (0.05 mM) added during the middle exponential phase growth (i.e., OD600 0.5) and at the same time decreasing the temperature to 30°C to minimize potential formation of inclusion bodies. After 12h, the induced-culture broth were centrifuged at 12,000 g for 30 min at 4°C.

1.14 $SHMT_{Sth}$ purification

All the further steps were carried out at 4°C.

The $SHMT_{Sth}$ cell pellet was re-suspended in 50 mL of lysis buffer A, containing 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, pH 8. Cells were lysed using a Cell Disrupter (Constant Systems) at 1.35 Kbar. Cellular debris was removed by centrifugation at 25,000 g for 20 min. The clear supernatant was collected, centrifuged at 13,000 rpm for 20 minutes (Sigma 2 16-PK) and subsequently filtrated. The filtrated sample is then loaded and purified by affinity chromatography in an FPLC system (Amersham biosciences), using a Nickel Sepharose High Performance (GE Healthcare) containing HisTrap chelating stationary phase (Amersham Biosciences) with 5 mL bed volume and then by Size Exclusion Chromatography (SEC) in the same system, using a Superdex-200 gel filtration column (GE Healthcare).

1.14.1 Nickel Affinity chromatography step

First a washing step of 5 column volumes with distilled water was performed. Then the column was equilibrated with lysis buffer A (5 column volumes at a flow rate of 3 mil/min. The crude filtered

supernatant ([12 mg/mL]) was applied to the column (divided in two runs) and subsequently washed with 25 mL of buffer B, containing 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole pH 8, at a flow rate of 1 mL mL/min. Finally the protein was eluted with buffer C, containing 50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole pH 8, at a flow rate of 1 mL/min.

Two fractions of 5 ml, corresponding to the elution pick were collected for the run number 1 and two fractions of 3 mL for the run number 2.

The eluted fractions were analysed by SDS-PAGE using 15% polyacrylamide gels. The purest fractions were selected, pooled and concentrated up to 10 mg/mL in a total volume of 2 mL, using microfiltrators (Vivaspin 20) with a cutoff of 20,000 Daltons. The concentration measured by nanodrop (Nanodrop 2000 by Thermo Fisher Scientific). The absence of nucleic acids is assessed by the O.D. relation at 280 and 260 nm (OD260/OD280) of the peak fractions, which must be below 0.7.

1.14.2 Size exclusion chromatography (SEC)

The concentrated fractions were filtered (Millex-GP, 0.22 μm) and further purified on a Superdex-200 gel filtration column (GE Healthcare) equilibrated with buffer C. The protein was eluted using the same buffer at a flow rate of 0.5 mL/min. Peak fractions corresponding to the dimer were pooled and concentrated up to 20 mg/mL using a 5,000 Daltons cutoff VIVASPIN. The protein was, finally, dialysed against HEPES 10 mM pH 7.5 and it was frozen at −80°C and lyophilized.

1.15 $SHMT_{Sth}$ crystallization

$SHMT_{Sth}$ *wild-type* protein (20 mg/ml) in a buffer HEPES (20 mM, pH 7.5) rendered monoclinic crystals in sodium citrate (0.8–1.2 M) and cacodylate (0.1 M, pH 6.5), using the sitting drop vapour-diffusion method at 293 K. The mixture, consisted of equal volumes of protein with precipitant solution (2 µL), was equilibrated against 500 mL of the reservoir solution. The crystals, which appeared after five days, were soaked with a cryo-protectant solution (1.2 M sodium citrate and 20–25% glycerol), before flash cooling to 100 K. The complexes of $SHMT_{Sth}$ *wild-type* with the cofactor PLP and Gly or L-Thr substrates were crystallized in a tetragonal form under similar conditions as the *apo*-$SHMT_{Sth}$ *wild-type*. The $SHMT_{Sth}$ *wild-type* in complex with glycine was crystallized from a mixture of Gly (17 mM), PLP (5 mM), cacodylate (0.1 M, pH 6.5) and sodium citrate (1.1 M). In the case of the L-Thr complex, 2 µL of the protein solution were mixed with 2 µL of reservoir solution, which contained additional PLP (0.1 mM) and L-Thr (100 mM), and together with cacodylate (0.1 M, pH 6.5) and sodium citrate (0.85 M). The crystals, which appeared between three and five days, were soaked with a cryo-protectant solution (1.2 M sodium citrate and 20–25% glycerol), before flash cooling to 100 K.

For both *apo*-form and complexes, the suitable crystallization conditions were found by extensive screening of chemical conditions employing automated dispensing methods. For this, 96 wells plates (Hampton Research) were dispensed by a nanodrop dispensing robot (Cartesian model by “Genomic solutions” or Phoenix by “Rigaku”). Nanodrops containing 100 nl of protein + 100 nl of crystallization buffer from the reservoir were prepared and monitored during time. The successful or promising crystallization conditions were then scaled

up to µL volumes using 24 well plates (Hampton Research) also using the sitting drop set up, as explained.

1.15.1 $SHMT_{Sth}$ Y55T micro-seeding experiment

For Y55T variant in complex with D-Ser micro-seeding experiments were performed in order to obtained suitable crystals for X-ray diffraction. The thin crystals appeared in a sitting drop, containing sodium citrate 1.2 M, cacodylate 0.1 M pH 6.5, 5 mM PLP and 100 mM D-Ser, were firstly crushed with a probe directly into the original drop, after adding 50 µL of reservoir solution to minimize and compensate for evaporation from the drop, and then removed and placed in a seed bead tube on ice.

450 µL of stabilizing solution (e.g. well buffer or something a little more concentrated than well buffer) were added to the seed bead tube (Hampton Researcher) and the tube was vortexed (3 min). Two dilutions of the stock (indicated as "dilution $1X10^{0}$") were prepared, adding 2.5 µL of the stock to 45 µL of stabilized solution containing no protein or crystals (dilution $1X10^{-2}$) or 2.5 µL of dilution $1X10^{-2}$ in 45 µL of stabilized solution containing no protein or crystals (dilution $1X10^{-4}$). Hanging drop plate were prepared using the same reservoir buffer used to obtain the initial seed crystals and putting 2 µL of seed solution (dilution $1X10^{-2}$ and $1X10^{-4}$) directly on the well. Crystal of suitable shape appeared in the drops prepared with dilution $1X10^{-4}$ between three and four days after. They were soaked with a cryo-protectant solution (1.2 M sodium citrate and 20–25% glycerol), before flash cooling to 100 K.

1.16 X-ray diffraction, data processing and structure determination

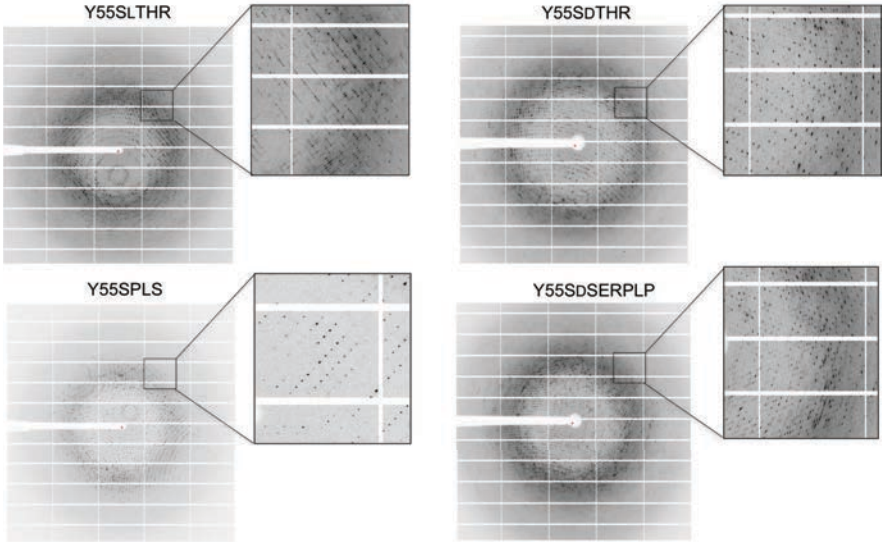
Three datasets of $SHMT_{Sth}$ wild-type were collected from the beamlines ID29 at the ESRF synchrotron in Grenoble, France. For the *apo*-$SHMT_{Sth}$ *wild-type*, a dataset with a total of 360 images was collected with an oscillation of 1° at 2.05 Å, using a charged-coupled (CCD) detector and a fixed X-ray wavelength at 0.976 Å. For the complexes of $SHMT_{Sth}$ *wild-type* with Gly and L-Thr, complete X-ray diffraction data sets of 180 and 180 images were collected with an oscillation of 0.5°, using a CCD detector and a fixed X-ray wavelength at 0.976 Å.

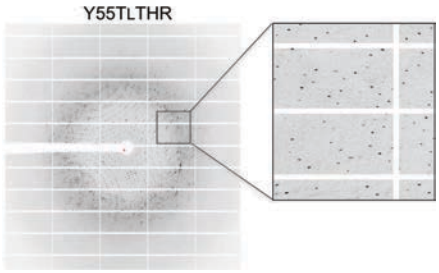
Four datasets of $SHMT_{Sth}$ Y55S and two of $SHMT_{Sth}$ Y55T were collected from the beamline XALOC of the ALBA Synchrotron in Barcelona, Spain. Five datasets with a total of 3600 images were collected with an oscillation of 0.1 at a resolution between 2.2 and 1.8 Å, using a PILATUS 6M detector and a fixed X-ray wavelength at 0.976 Å. In all the cases diffraction data were processed with XDS program (Kabsch, 2010). Intensities were scaled with SCALA (Evans, 2006) and converted to structure factor amplitudes using C-truncate (Collaborative Computational Project number 4, 1994), both programs from CCP4 suite (Winn et al., 2011).

The *apo*-structure was solved by the Molecular Replacement method (Read R. J. et al., 2007) using Phaser from the CCP4 suite (McCoy et al., 2007) and employing the structure of SHMT from *Bacillus Stearothermophilus* as a search model (PDB: 1KKJ) (Trivedi et al., 2002). The model was manually completed using COOT (Emsley and Cowtan, 2004) and was refined with Refmac5 (Murshudov et al., 1997).

The structure of the complexes was solved by Molecular Replacement, using Phaser from CCP4 suite and the native structure of *apo*-$SHMT_{Sth}$ a search model (PDB code: 4WXG). The model was manually completed using COOT (Emsley, 2004) and was refined with Refmac5 (Murshudov et al., 1997). The ligands were manually adjusted and built in the elctron density map. Solvent molecules were added during final cycles of refinement. The crystallographic free R-factor (Brunger, 1992) was monitored at each stage to prevent model bias. The quality of the structures was evaluated using the Ramachandran plot (the plots corresponding to each structure are shown in the picture above).

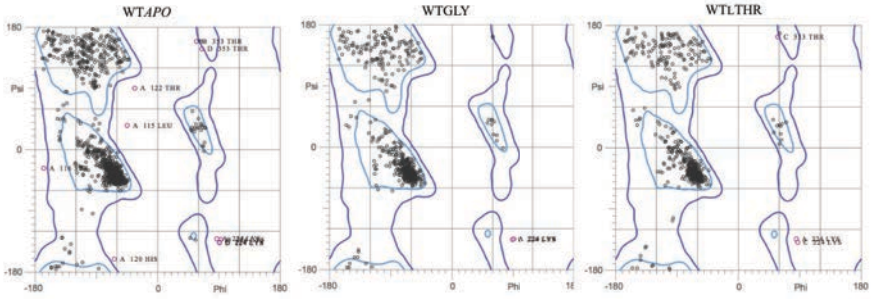
X-ray diffraction patterns of the variants



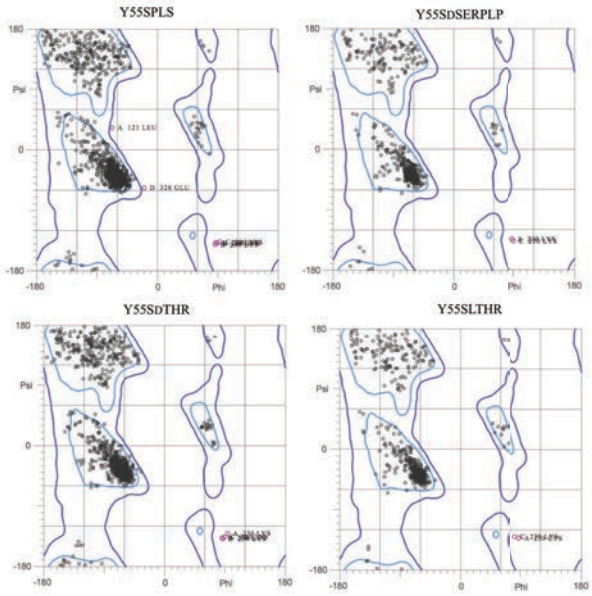


Ramachandran plots

Ramachandran plots showing the final refined model for SHMTSth *wild-type* and variants



Number of resiued in favoured region:		
1549 (96.33%)	786 (97.28%)	795 (97.91%)
Number of residue in allowed region:		
40 (2.49%)	18 (2.23%)	13 (1.60%)
Number of residues in outlier region:		
19 (1.18%)	4 (0.50%)	4 (0.49%)



Number of residues in favoured region:

794 (97.54%) 1565 (96.25%)

1572 (96.92%) 1550 (95.56%)

Number of residue in allowed region:

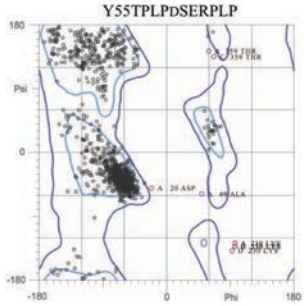
15 (1.84%) 51 (3.14%)

40 (2.47%) 58 (3.58%)

Number of residues in outlier region:

5 (0.61%) 10 (0.62%)

10 (0.62%) 14 (0.86)



Number of resiued in favoured region:
1550 (95.56%)
Number of residue in allowed region:
58 (3.58%)
Number of residues in outlier region:
14 (0.86%)

Conclusions

CHAPTER 1 Structure and catalytic mechanism elucidation of Serine hydroxymethyl transferase from *Streptococcus thermophilus* ($SHMT_{Sth}$) and variants.

I. The full-lenght $SHMT_{Sth}$ of 43 kDa was overexpressed, purified and crystallised in its *apo*-form and in complex with glycine and L-threonine (*halo*-forms).

II. The structures of *apo* and *halo*-$SHMT_{Sth}$ were solved by X-ray diffraction crystallography. The molecular replacement (MR) method was employed to solve the phase problem, using the structure of SHMT from *Bacillus stearothermophilus* as a search model. The corresponding PDB accession numbers are: 4WXB for the WT*apo*; 4WXF for WTgly and 4WXB for WTLThr.

III. The structural analysis of the catalytic active site in its atomic detail shed light on the catalytic cleft architecture and the amino acid species involved in the enzyme reaction and guided the protein engineering experiments on the *wild-type*.

IV. Through site-direct mutagenesis, different variants were produced and tested in our research group. $SHMT_{Sth}$ Y55S and Y55T variants appeared to be the best catalysts for the aldol addition of D-Ser. Both variants were produced, overexpressed and purified. Suitable protein crystals were produced for the X-ray diffraction analysis.

The Y55S variant was crystallized in complex with L and D-threonine and D-serine (external and internal aldimine complexes); the Y55T variant was crystallized in complex with L-serine (internal and external aldimine complexes).

V. Thanks to the structural analysis of the reaction intermediate obtained, we clarified important aspects of the enzyme catalytic mechanism and we proposed that the base responsible for the proton abstraction in the aldol addition reaction is a crystallographic water molecule, which appear to be activated by H bonds to the PLP-phosphate and Glu-57 and His-126.

VI. D-serine is not a substrate for *wild-type* enzyme, whereas it is a good substrate for the variants. The D-serine external aldimine found in the Y55SPLS and Y55TDSerPLP structures, shows that the empty space generated by the mutation in position 55 is partially occupied by the hydroxymethyl group of D-serine, which is now accessible to the catalytic water for the proton abstraction. In the *wild-type* structure, the D-Ser substrate adopts an opposed orientation, hindering the access to the H-Cα atom, which would be at more than 5 Å from the Glu-57 and the catalytic water molecule, as suggested by computational modelling performed in our research group.

VII. In the Y55T variant, the H-Cα is closer to residue Glu-57 (2.8 and 3.2 Å in the Y55SPLS and Y55TDser structures respectively) and to the water molecule (2.9 and 3.5 Å), resulting in a more efficient proton transfer and enhanced catalytic reaction of this variant.

VIII. The mutation in position 55 affects the LTA activity of the enzyme. Examination of the active site of $SHMT_{Sth}$ Y55S and Y55T variants shows that the space generated by the substitution of the tyrosine residue determines a reorientation of the hydroxyl group of the L-threonine substrate in the external aldimine complex, which will be now more distant from the Glu-57 (2.9 and 3 Å) and from the catalytic water molecule. This affords an explanation why the absence of retroaldol activity of Y55X variants was detected against L-threonine: the cleavage of the external aldimine to quinonoid complex is not possible anymore and the reaction is stuck in this step, contrarily to that observed for the SHMT *wild type*.

IX. $SHMT_{Sth}$ variants are not active with D-threonine substrate. Analysing the X-ray structure of Y55S in complex with D-Thr, we demonstrated that D-Thr displaces the PLP, at the position normally occupied by its phosphorilated group. Simultaneous closure of the active site stabilizes the inactivated enzyme.

X. The analysis of two consecutive steps of reaction, D-Ser internal aldimine and external aldimine, reveals that the PLP ring undergoes a 25 degree rotation, as observed in other SHMTs from different organism. Moreover, the observation that the catalytic lid is closed in both moments, suggested that conformational change happens when the substrates are both in the active site and not just on the verge of catalytic reaction.

XI. The structural studies of this interesting biocatalyst and its variants, in order to elucidate their catalytic mechanism, offered a rational basis to design improved or customized catalytic properties. Such knowledge could provide the basis for new medical treatments, pollution-control strategies, fuel production, among many other applications.

CHAPTER II

Structural solution of halohydrin dehalogenases D2 from *Gammaproteobacterium*

Introduction

2.1. Halohydrin dehalogenases

Halohydrin dehalogenases (HHDHs) are lyases that catalyze the cleavage of carbon-halogen bond of halohydrins. They also can catalyze the reverse reaction in the presence of nucleophiles such as cyanide, azide and nitrite ions in an irreversible epoxide ring opening reaction, resulting in the formation of novel C-C C-C, C-N or C-S bonds.

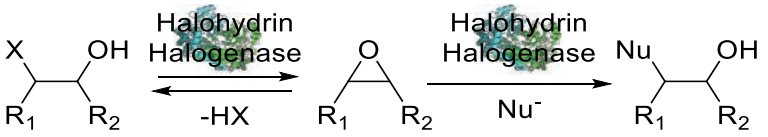


Figure 33: HHDHs catalytic reaction

Halohydrin dehalogenases are enzymes that convert vicinal halohydrins to an epoxide and a halide ione. They are also able to catalyse in the presence of a nucleophic species, the irreversible epoxide ring opening reaction.

In 1968, Castro and Bartnicki reported for the first time an enzyme, isolated from *Flavobacterium* sp., involved in the degradation of 2,3-dibromo-1-propanol by removal of the bromo-substituent in 2-position and concomitant formation of the corresponding epoxide epibromohydrin (Castro and Bartnicki, 1968). The enzyme was then called halohydrin epoxidase and it is a member of the dehalogenase enzyme group (Janssen, 2007). Since then, a number of enzymes able to dehalogenate various vicinal halohydrins have been isolated from different bacterial species and their biochemical and biocatalytic properties have been characterized. These enzymes, also called halohydrin dehalogenases (HHDHs), belong to the enzyme class of lyases and do not require any cofactors for keeping their catalytic

activity. The reaction catalysed that results in the formation of novel C-C, C-N, or C-O bonds, and nucleophilic promiscuity of HHDHs, together with the enantioselectivity that some of them show in their catalytic processes, makes them potential candidates for practical application in biocatalysis.

HHDHs usually appear in microorganisms, which utilize halohydrins as carbon and energy sources. They remove halogen substituents in order to detoxify the compounds and allow them to enter central metabolism. During the biodegradation of epichlorohydrin, for example, the compound is first ring-opened by an epoxide hydrolase and then dehalogenated to form glycidol, an epoxide (Arjan et al., 1989).

Despite the large number of studies dealing with the isolation of microbial strains that exhibit HHDH activity and purification of the respective enzymes, only very few of the corresponding HHDH-encoding genes have been cloned.

For this reason, until now, only five different HHDH sequences have been published and assigned to one of the three different phylogenetic subtypes A, B, and C, based on the sequence homology: HheA from *Corynebacterium* sp. strain N-1074 (Yu et al., 1994), HheA2 from *Arthrobacter* sp. strain AD2 (van den Wijngaard et al., 1991), HheB from *Corynebacterium* sp. strain N-1074 (Yu et al., 1994), HheB2 from *Mycobacterium* sp. strain GP1 (van Hylckama Vlieg et al., 2001), and two identical HheC sequences from *Agrobacterium radiobacter* AD1 (van Hylckama Vlieg et al., 2001) and *Rhizobium* sp. strain NHG3 (Higgins et al., 2005). Of these, HheC from *A. radiobacter* AD1 has been the most studied, as this enzyme seems to show an exceptional enantioselectivity (Koopmeiners et al., 2016).

Sequence analysis indicates that HHDHs are very close to each other in the same group, with an identity of 88.7-98.3%; however, they have significant differences among different groups, where identity falls to 18.9-33.5 %.

2.2. HHDHs, structures and catalytic mechanism

Phylogenetically, HHDHs belong to the superfamily of short-chain dehydrogenases/reductases (SDRs), with which they share several structural and mechanistic features (van Hylckama Vlieg et al., 2001) (Kavanagh et al., 2008).

The first 3D structure of a HHDH, HheC from *A. radiobacter* strain AD1, released in 2001 was an homology model derived using the coordinates of SDR family members as template (Vlieg et al., 2001) and the structural studies aimed to predict the structural basis of the catalytic reaction.

Subsequently, the X-ray structure of HheC was solved at 3.0 Å resolution (de Jong et al., 2006). To date, in the PDB are available also the crystal structures of HheA (Watanabe et al., 2015), HheA2 (de Jong et al., 2006), HheB (Watanabe et al., 2015), and HheC (de Jong et al., 2006) and they reveal an overall highly similar structural fold.

HHDHs, studied until now, have a size range of 20-35 KDa. They are homotetramers formed by a pair of dimers, and possess a catalytic triad made of Ser-Tyr-Arg in the C-terminal part, differing from members of SDR family, which show the presence of a lysine residue instead of the arginine in the triad. A spacious nucleophile binding pocket is present at the same position in which a nicotinamide cofactor binding site (NAD(P)H-binding site) has been found in the enzymes belonging to

the SDR family. The highest degree of conservation between members of the SDR protein family is observed within the N-terminal part of these proteins, which contains a typical (G/A)-(G/A)-X-X-(G/A)-X-G fingerprint. This fingerprint is characteristic for Rossmann fold of the cofactor binding site (Jornvall et al., 1995), but it is not conserved in the HHDHs. The absence of a cofactor-binding motif in halohydrin dehalogenases is in agreement with the fact that the dehalogenation reaction that is catalysed by these enzymes is not a redox reaction. However, the HHDHs monomers exhibit the typical Rossmann fold, also found for other SDR enzymes, consisting of a six- or seven-stranded parallel β-sheet that is surrounded by seven or eight α-helices.

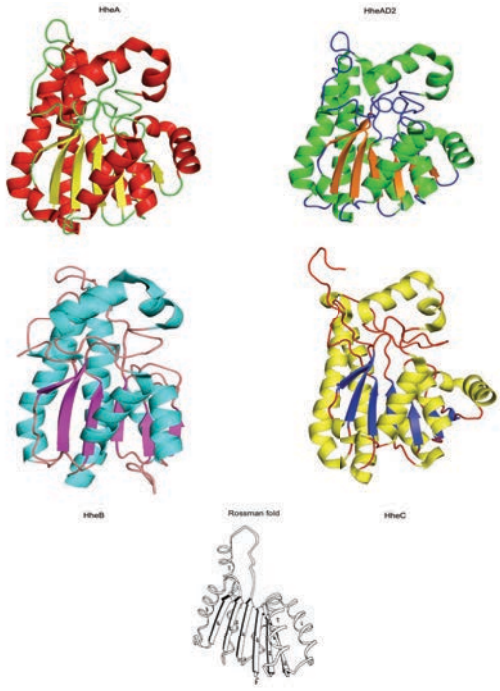


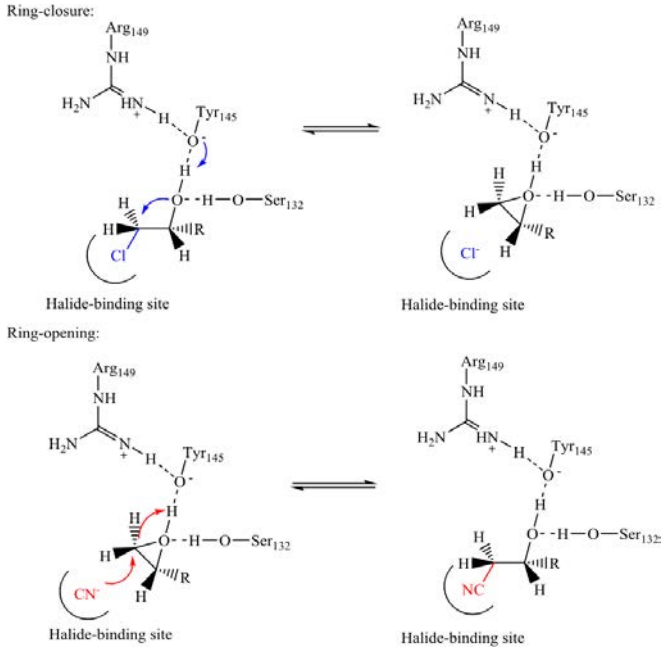
Figure 34: Three-dimensional structures of available HHDHs

3D-structures of the monomer of HHDHs available in the PDB up to now, showing the typical Rossman fold. From left hand side to right hand side: HheA, HheA2, HheB and HheC.

Each monomer active site is buried within the enzyme and is connected with the bulk solvent via a substrate entrance tunnel. This tunnel is differently shaped in the crystallized HHDHs and, at least for HheC, various studies demonstrate that residues of the entrance tunnel are involved in determining the enzymes activity and enantioselectivity (Schallmey et al., 2013) (Guo et al., 2015) (Wang et al., 2015). The position of the catalytic triad, formed by residues Ser, Tyr, and Arg, which are present within the active sites, is also highly conserved.

The catalytic process under HHDH is proposed to divide into "ring closure" and "ring opening" steps. During dehalogenation, the Ser residue positions the haloalcohol substrate within the active site by formation of a hydrogen bond with the substrate hydroxyl group. The Tyr, as the catalytic base, takes up a proton from that hydroxyl group, adjacent to the halogen atom and concurrently the substrate oxygen performs a nucleophilic attack on the halogen-bearing carbon atom, resulting in the ring closure and release of a halogen. The Arg residue does not interact with the substrate itself but lowers the pKa of the tyrosine-OH via hydrogen bonding in order to activate it for proton abstraction (de Jong et al., 2006). In the reverse reaction, epoxide ring opening, the Tyr donates a proton to the forming oxygen anion after nucleophilic attack of a halide (or another accepted nucleophile) on one of the carbons of the epoxide ring.

The nucleophile binding pocket in HHDHs is formed by residues mapping to the glycine-rich T-G-X3-A/G-X-G motif of SDR enzymes as well as residues from other loops forming hydrogen bonds from backbone amides with the anionic nucleophile (de Jong et al., 2006). Additionally, a water molecule present in close proximity supports the hydrogen bond network around the nucleophile binding pocket.



Scheme 5: The catalytic mechanism of halohydrin dehalogenases

Reaction mechanism and active site residues of HHDHs, exemplified by halohydrin dehalogenase C (HheC) from *Agrobacterium radiobacter* AD1. Interactions of the conserved amino acids of the triad with halohaydrin are shown. Catalytic mechanism with deprotonation of the substrate hydroxy group by Tyr-145, assisted by Ser-132 and Arg-149.

2.3. Halohydrin dehalogenases biocatalytic applications

dehalogenases are remarkable enzymes with an exceptional potential for biocatalysis, since they are able to form novel C-C, C-O, C-N, or C-S bonds. Moreover they are promiscuous enzymes, with epoxide ring-opening activity, able to react with a wide range of nucleophiles. The HHDH-catalyzed stereoselective ring opening of epoxides is widely used in the asymmetric synthesis of enantiopure β-substituted

alcohols, which are building blocks of pharmaceuticals and biologically active compounds (Elenkov et al., 2006).

HHHD has also an important role in the biodegradation of halogenated environmental pollutants. Halogenated organic compounds (HOCs) represent a large class of natural synthetic substances that contain one or more halogens combined with carbon. They are widely used in the production of insecticides, fungicides, herbicides, intermediates of pharmaceuticals, etc. With their wide range of applications, HOCs inevitably enter into the environment and unfortunately they are toxic and harmful for human health and classified as environmental pollutants (Emanuelsson et al., 2008) (Olaniran et al., 2004). Moreover, due to their chemical stability, they do not decompose spontaneously. The biodegradation of these compounds has received considerable attention, because a reliable and efficient decontamination method is required. It is reported that HHDHs play a major rule in the degradation of HOCs (Janssen, 2007), thanks to their ability to remove the halogen atom from these compounds. For example, HHDHs are able to degrade halohydrin, 3-choloro-1,2-propanediol (3-CPD), and its fatty acid esters occurred in food recently (Hamlet et al., 2002) and their removal is necessary to prevent health issues. Bornscheuer and co-workers have reported an enzymatic process to remove 3-CPD, which was converted into the harmless product glycerol by HHDH and epoxide hydrolase (Bornscheuer and Hesseler, 2010).

2.4 Halohydrin dehalogenase from *Gammaproteobacterium*

Among the pool of known HHDHs, only a limited number of catalytic applications have been so far explored, whereas their potential for organic synthesis remains largely unexploited. Indeed, after the identification of specific sequence motifs, 37 novel enzyme sequences were readily identified in public sequence databases (Schallmey et al., 2014). Phylogenetic inference for enzymes of the halohydrin dehalogenase family confirmed that they form a distinct monophyletic clade within the SDR superfamily. In addition, the majority of these novel enzymes were substantially different from previously known HHDH subtypes. Consequently, four additional phylogenetic subtypes were defined, greatly expanding the halohydrin dehalogenase enzyme family.

In the group of Prof. Annett Schallmey from the Institute for Biochemistry, Biotechnology and Bioinformatics in Braunschweig (Germany), which is also involved in the “Carbazymes” European Project, the novel HHDHs are being tested for different synthetic scopes. In particular, HHDH´s potential to convert the deoxyribose phosphate aldolase (DERA) products in chiral epoxide, which are valuable and essential building blocks for the synthesis of enathiomerically pure pharmaceuticals and agrochemicals (Choi et al., 2000), was tested. HheD2 was the most promising candidate, and this initiated our cooperation for the structural study in order to characterize the protein and guide further rational design.

Results and Discussion

2.5 HheD2 purification

HheD2 was produced in the group of Prof. Annett Schallmey from the Institute for Biochemistry, Biotechnology and Bioinformatics in Braunschweig, Germany. Our collaborators sent us a cell pellet, containing 1mg HheD2 for the crystallization experiment, X-ray diffraction and structural determination. In our laboratory, we established optimal conditions for protein solubilisation during cell lysis, performed in a french press step using a cell disrupter on the bacterial cells in a lysis buffer that contained 50 mM HEPES ph8, 500 mM NaCl, 25 mM imidazole, 1 mM DTT, DNAse, and protease inhibitors. After supernatant clearance by centrifugation, a two-step purification protocol followed. First the filtrated supernatant was passed through a nickel-chelating column that traps histidine-tagged proteins, which are eluted by a gradient of imidazole. The purest eluting fractions were pooled, concentrated and injected to a gel filtration column. A single, pure band of HheD2 was obtained. It migrated as a molecule of 26 kDa after the second purification step. The corresponding fractions were pooled, concentrated and dialysed against HEPES 50 mM. The protein was obtained with a level of purity higher than 95% as assessed by SDS-PAGE (**Fig.35**). The final yield obtained, after protein purification, was 13 mg of HheD2 in 800 µL.

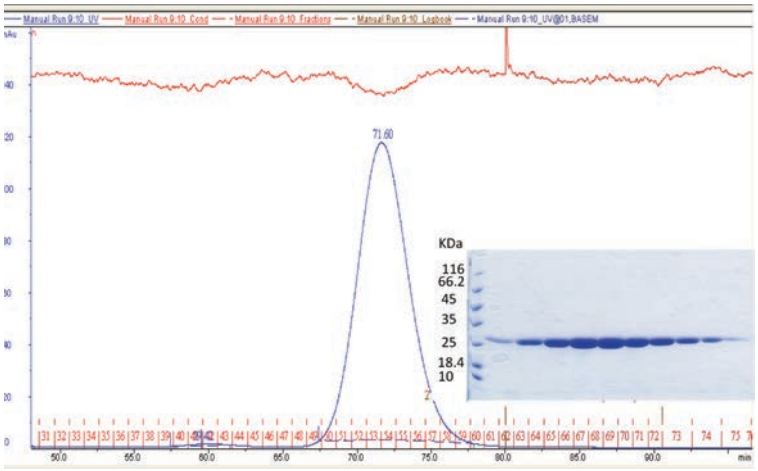


Figure 35: HheD2 purification chromatogram and SDS-gel.

HheD2 was purified with a Superdex 200 16-60 column (GE Healthcare). The absorbance at 280 nm is shown as blue line. The eluted fractions from the peak were loaded into a poly-acrylamide denaturating gel (on the right hand side of the picture). A single pure band of the protein was obtained that migrates as a molecule of 26 kDa.

2.6 HheD2 crystallization and data collection and processing

In order to elucidate the tridimensional structure of HheD2 in atomic detail, crystallization experiments were performed and suitable crystals for the X-ray diffraction were obtained in two different crystallization buffers.

Bipyramidal crystals appeared after three days in a crystal buffer (indicated here as “crystal buffer I”), containing lithium sulfate 0.2 M, CHES 0.1 M pH 9.5, and Na/K tartrate.

Prismatic crystals appeared after seven days in a crystal buffer (indicated here as “crystal buffer II”), containing sodium acetate 0.2M, Tris 0.1M pH 8,5, 30% PEG 4000 (**Fig.36**).

The protein crystallizes in two different crystal forms. Protein crystals with space group $P2_12_12_1$, and unit-cell parameters a=76.2, b=93.9, c=140.5, $\alpha=\beta=\gamma$=90.0º, were obtained for the protein crystallised in crystal buffer I. Space group $P2_12_12$, with unit-cell parameters a=120.44 Å, b=130.99 Å, c=140.65 Å, $\alpha=\beta=\gamma$=90.0º, were obtained for the protein crystallised in crystal buffer II.

Both crystals diffracted to high resolution: in the case of $P2_12_12_1$ space group a resolution of 1.9 Å resolution was obtained; and an even better resolution, 1.8 Å, was obtained for $P2_12_12$ space group.

For both structures a dataset with a total of 3600 images spanning a full rotation of 360º was collected from the beamline XALOC at the ALBA synchrotron (Barcelona, Spain), using a PILATUS M6 detector at a fixed wavelength.

For convenience, the protein crystals belonging to $P2_12_12_1$ space group and the corresponding structure will be indicated, from now on, as "**crys_I**", referring to the corresponding crystallization buffer; the protein crystals with P21212 space group and their corresponding structure will be indicated as "**crys_II**".

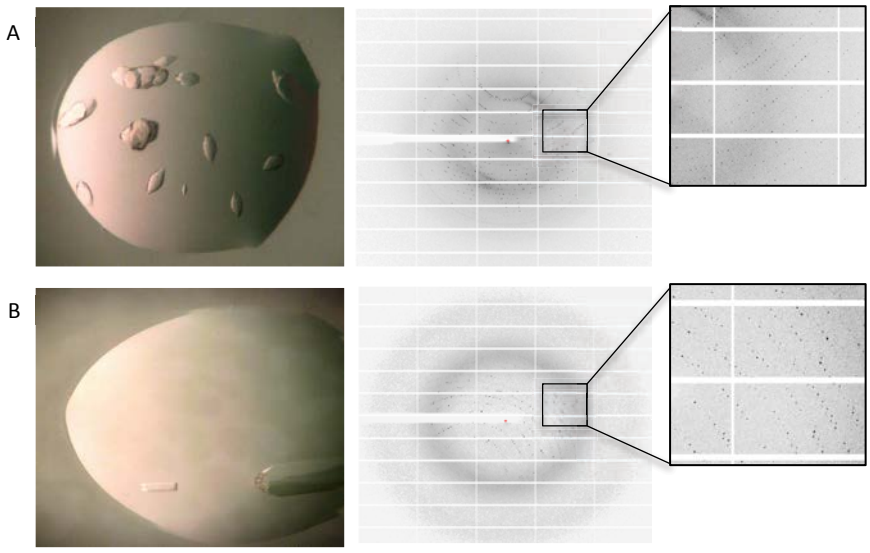


Figure 36: Crystals of CrysI and CrysII and samples of their corresponding X-ray diffraction patterns.

Images of the crystallization drops and X-ray diffraction patterns of typical crystals of HheD2. (**A**) Bipyramidal crystals appeared in buffer I and corresponding and the best dataset diffracting to 1.9 Å; (**B**) Prismatic crystals appearing in buffer II and the best dataset diffracting to 1.8 Å resolution.

The calculated Matthews coefficient (Matthews, 1968) for CrysI was of 2.52 $Å^3$/Da, which is consistent with four monomers of the protein in the asymmetric unit. In fact, the structure contains a biologically functional tetramer of HheD2 per asymmetric unit. The calculated Matthews coefficient for CrysII was of 2.77 $Å^3$/Da, which is consistent with eight copies in the asymmetric unit.

Diffraction data and refinement statistics of the two structures are summarised in **Table 5.**

	Crys_I	**Crys_II**
Data collection		
Source	XALOC	XALOC
Wavelenght (Å)	0.9699	0.9330
Space group	$P2_12_12_1$	$P2_12_12$
Cell dimensions		
a, *b*, *c* (Å)	76.2-93.9-140.5	120.4-130.9-140.6
a, *β*, *γ* (°)	90.0-90.0-90.0	90.0-90.0-90.0
A.U. content	one tetramer	two tetramers
Resolution (Å)	78.13–1.9(1.84-1.90)	95.8-1.8 (1.86-1.89)
Unique reflections	133216 (15884)	107511 (7867)
Completeness (%)	95.4(45.8)	99.9 (99.9)
Redundancy	3.6 (3.6)	8.1 (8.0)
I/σI Redundancy	10.3 (2.2)	14.4 (2.2)
Refinement		
Resolution (Å)	78.1 – 1.84	95.86 – 2.2
No. reflections	87632	107511
R_{work} / R_{free}	0.18/ 0.20	0.24/0.30
No. atoms		
Protein	6975	13074
Ligand	58	0
Water	802	629
B factors		
Protein	31.94	39.7
Water	40.80	33.5
Ligand/ion	36.34	
R.m.s. deviations		
Bond lengths (Å)	0.012	0.012
Bond angles (°)	1.18	1.64

Table 3: Crystallographic statistics on data collection and refinement

Values within parentheses correspond to the last resolution shell.

2.7 HheD2 structure solution using ARCIMBOLDO-SHREDDER

Molecular replacement relies on the availability of suitable models to be placed in the unit cell of the unknown structure in order to provide initial phases. In order to phase HheD2, a search of suitable models was performed by using HHPRED (Soding et al., 2005). This algorithm uses, apart from the identity between aligned sequences, information about the match between predicted secondary structure in the input sequence and the one for the structures in the database. This allows finding more distant homologs that might be less related in sequence but still preserve fold similarity.

From the pool of homologs three were selected: HhedB (PDB: 4ZD6) with a sequence identity of 47%; HheA (PDB: 4Z9F) with a sequence identity of 30%; Hped, (PDB: 4URF) with a sequence identity of 25%. Our group developed a program, ARCIMBOLDO_SHREDDER that derives fragments from an initial distant homolog template in order to use them to phase the structure. Therefore, the three models were used as input template for SHREDDER. A common parameterisation was used in the runs. Starting from the initial model, coil regions were removed and only α-helices composed by more than 7 residues and β-strands larger than 4 residues were left. The template model was then analysed in order to predetermine the choices in the generation of fragments, as well as their progressive decomposition in different rigid body groups for subsequent refinement steps. This refinement is accomplished in decomposition of the secondary structure elements into a few groups defined by distance and preservation of folds such as association of strands into a sheet. A second level further disengages individual helices (**Fig.37**).

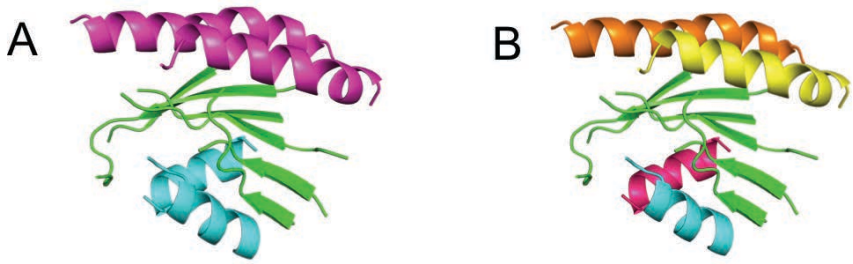


Figure 37: Levels of model decomposition in the refinement process

Rigid body progressive decomposition in secondary structure elements used for refinement steps. In (**A**) the helices with the same colour represent two groups of rigid body secondary structure elements. The β-sheet represents another rigid body group. In (**B**) the helices are further decomposed into four groups of rigid body refinement.

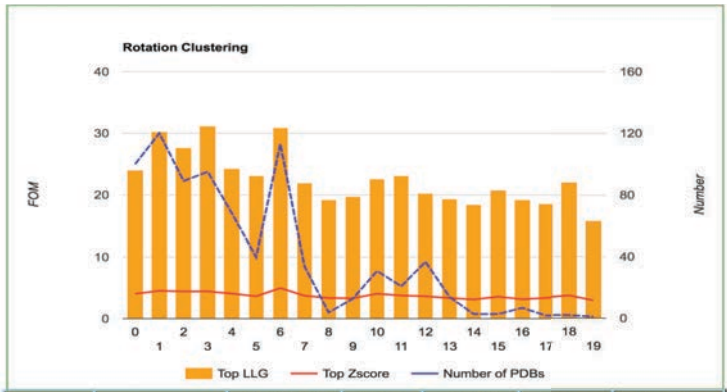
At this stage, the template is used to generate up to 101 non-redundant models, by using a distance-based approach that produces compact models. The partition scheme established on the template is adopted for each of the partial models derived.

These models are evaluated as a library using ARCIMBOLDO_BORGES (Millan et al., 2015). In all runs, rotation search and first cycle of gyre refinement was performed at 0.8 Å r.m.s.d.. A second cycle of gyre refinement and the rest of the PHASER steps were performed at 0.5 Å r.m.s.d..

Correctly placed solutions were found for both rotation clusters 6 and 3 (**Fig.39**), showing Z-scores over 7 that survive the packing check and discriminating values of LLGs, after rigid body refinement and gimble. However these individual solutions, after density modification and autotracing with shelxe, did not develop to a complete solution. Arcimboldo_borges searchs for just one copy of the models in the library, so different rotation clusters might match different copies of the asymmetric unit. At this stage Alixe (Millán et al., 2015a) was used

to combine phase sets from the different rotation clusters. The resulting combinations were subjected to density modification and autotracing. A successful expansion, reaching a correlation coefficient (CC) of 33.58% was achieved with cluster joining up to 107 phase sets coming from 2 rotation clusters

A



#Cluster	#Rotations ▲	#Distinct Pdb	Top LLG	LLG Mean	Top Zscore	Zscore Mean
1	120	120	30.20	22.23	4.52	3.70
6	113	113	30.87	23.85	4.96	3.92
0	100	100	24.06	21.05	4.04	3.55
3	95	95	31.14	21.66	4.43	3.61
2	89	89	27.66	22.27	4.41	3.71
4	68	68	24.24	19.75	4.07	3.38
5	40	40	23.06	18.33	3.64	3.19

B

#Cluster	Rotation and Model Refinement					Translation Function					Packing					Rigid Body Refinement		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG
1	240	97.00	46.19	4.52	1.85	240	100.10	58.14	6.08	5.13	240	100.10	58.14	6.08	5.13	240	113.90	66.79
3	190	87.00	46.05	4.43	1.80	190	98.37	62.28	6.41	5.48	190	98.37	62.28	6.41	5.48	190	130.80	73.52
2	178	100.00	50.99	4.41	1.86	178	101.19	62.92	6.09	5.10	178	101.19	62.92	6.09	5.10	178	122.30	72.15
6	226	98.00	50.53	4.96	1.96	226	113.25	68.89	7.83	5.60	226	113.25	68.89	7.83	5.60	226	137.20	81.42

Figure 38: Html outputs from successful Arcimboldo_Shredder run

(**A**) Graph and table describing the rotation cluster. (**B**) Table with figures of merit for all phaser steps. Each row corresponds to the evaluation of a rotation cluster.

2.8 Overall structure of HheD2

The final model of HheD2 CrysI contains a biological tetramer in the asymmetric unit, constituted by four chains (A-B-C-D) of 223 residues (23-246), four molecules of CHES (distributed in the four active sites) coming from the crystallization buffer, and 802 water molecules. Electron density is clear and continuous from amino acids 11 to 246 in all the monomers. A complete structural model could be built, and only thirteen residues (11-24) were missing. The histidines 4-11 of His-tag, present at the N-terminal end due to the purification strategy, were also trace in monomer A.

In the final model of HHeD2 CrysII two tetramers were found (chains A to H) in the asymmetric unit, with each monomer constituted by 223 residues (23 to 246) and 629 water molecules. No electron density corresponding to residues 11-23 and the attached N-terminal His-tag, presumably disordered, can be seen; whereas, the remaining polypeptide chain is well defined in the electron density map and only eight residues (184-206) in chains C and D, contributing to the catalytic pocket, as explained later in this session, appear to be disordered and, consequently, are not included in the final model.

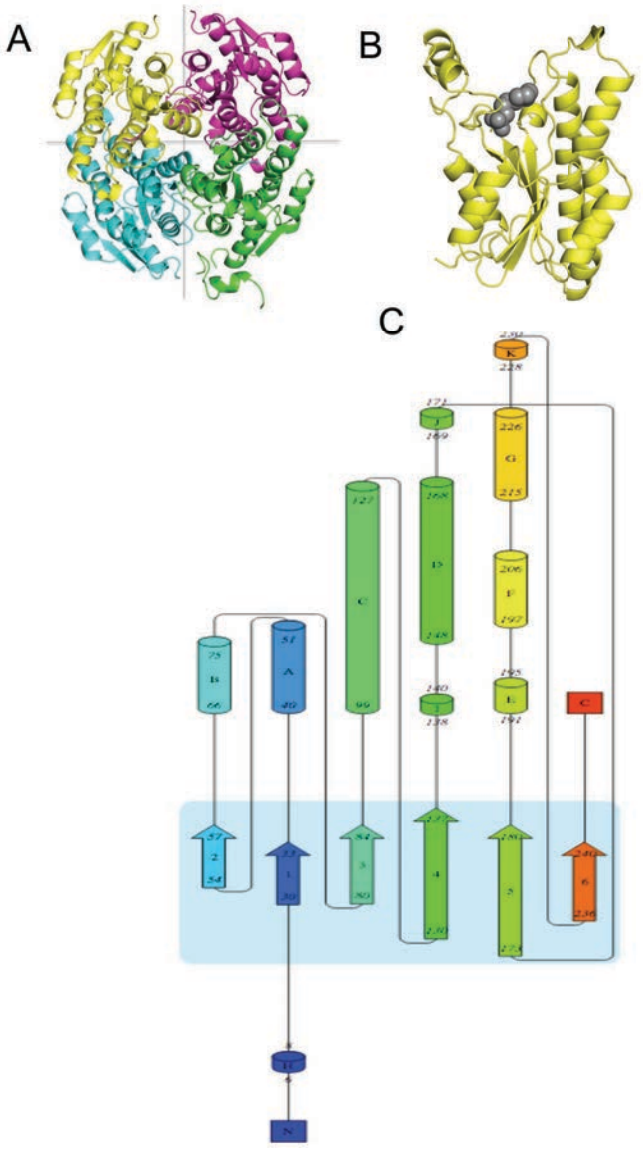
The protein of 26.46 KDa is a tetramer in solution, as demonstrated by the analytical size exclusion chromatography performed. Also the PISA program (Krissinel, 2015), which evaluates the energy of protein-protein interfaces observed in a packing structure and predicts which of these contacts would be stable in solution, confirms that the protein is a tetramer with 222 point group symmetry that can be regarded as a dimer of dimers (**Fig.39-A**). The fold and oligomerization interfaces of HheD2 are typical for members of the

SDR family, which mainly occur as tetramers or dimers (Jornvall et al., 1995).

The different subunits of the tetramer superimpose with r.m.s.d.s. between 0.11 and 0.19 for 246 Cα atoms, suggesting they are equivalent to each other.

The stereochemistry of the refined model is good; the Ramachandran plot shows that 96% of all residues fall into the core region as defined by Kleywegt and Jones (Kleywegt and Brunger, 1996) and two or seven non-glycine, non terminal amino acid residues lie within the core or most restrictive extension region of allowed values for Crys_I and Crys_II respectively. For both structures the outliers regard amino acidic residues present at the active site. This can be explained considering the flexibility required by the active site to undergo changes during catalysis and stabilize the reaction intermediates. The Ramachandran plots corresponding to the two structures are shown in the Experimental Procedure session.

In the HheD2 monomer six β-strands form a large parallel β-sheet structure, interacting through hydrophobic and electrostatic interactions of their side chains. A total of seven α-helices are distributed on both sides of the β-sheet, which resembles the characteristic Rossman fold. Two monomers form a dimer, in which the longest α-helices, αC and αD, and part of their connecting loops associate into an intermolecular, anti-parallel, four-helix bundle.



F

igure 39: Three-dimensional structure of HheD2

(**A**) Cartoon representation of the tetrameric structure of HheD2. The dimer interactions are formed by two pairs of α-helices that form a four-helix bundle in the central horizontal plane. The tetramer interactions are formed by the dimers packing along the vertical plane. (**B**) Cartoon view of the HheD2 monomer. In gray is shown the CHES molecule in order to highlight the active site position. (**C**) HheD2 econdary structure topology diagram.

Although the amino acid sequence identity of HheD2 with HheA, HheAAD2, HheB and HheC is relatively low (30%, 47%, 22%) respectively), the tertiary structure of HheD2 is very similar to that of these three proteins. When superposed the five monomeric structures show an r.m.s.d. of 1.22 Å, 1.21 Å, 0.70 Å and 1.31 Å respectively for all the Cα atoms (**Fig.40**).

One of the most remarkable differences is the lack of additional secondary structure elements: HheD2, as HheAD2 and HheB, lacks a β-strand and a α-helix with respect to HheC. Thus the loop of Asp-59 to Ala-66 has connected to β2 and αB. These regions do not contribute to the tetramer formation because they are surface-exposed regions and are not located in the core of the structure.

Another significant difference in HheD2 is the absence of the C-terminal extension that donates a tryptophan residue (Trp-249) to the active site of an adjacent subunit in HheC. The C-terminal end of each HheC monomer binds the 2-fold symmetry-related subunit that lies opposite in the other dimer (de Jong et al., 2006). The C-terminal extension is absent also in the HheA, HheAD2 and HheB structures. (**Fig.40**). The relevance of this extension is not known, but site-direct mutagenesis demonstrate that has effect on activity and selectivity (Tang et al., 2003). In fact, complete truncation of this C-terminal overhang (residues 246–254 in HheC) resulted in an overall lower thermostability as well as drastically reduced activity. The subtypes A- and B-type dehalogenases have only a modest enantiopreference (van den Wijngaard et al., 1991), showing marked differences in their catalytic behaviours with respect to HheC.

Other structural differences are present in the substrate binding pocket and the halide-binding site and are commented below.

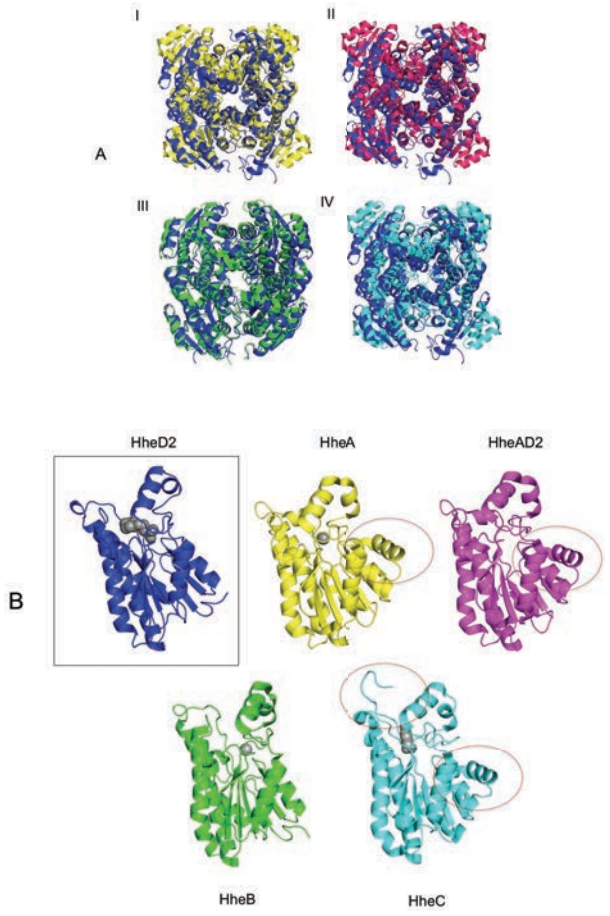


Figure 40: Structural comparison between HHDHs structures available in the PDB

(**A**) Superposition of the HheD2 structure (in blue) with (I) HheA (in yellow); (II) HheAD2 (in magenta) (III) HheB (in green); (IV) HheC (in cyan) reveals a similar overall structure organization. (**B**) Comparison of the monomeric structure of HheD2 with HheA, HheB and HheC (from the left side to the right side). The monomers have the same colours of the entire structure in the (A) and the active sites are identified by the ligands shown in sphere representation in gray. The main differences, regarding the C-terminal extension and the absence of an α-helix and a β-strand, are indicated by spotted red circles.

2.9 HheD2 active site architecture

The halide binding of HheD2 is formed by two loops extending from βαβαβ Rossman fold motif. The first and largest one is equivalent to highly variable loop that is responsible for cofactor and substrate binding in members of the evolutionary related SDR family, whereas the second loop corresponds to the Gly-rich cofactor-binding motif of the SDR enzyme (de Jong et al., 2003) (Oppermann et al., 2003).

In HheD2 Crys_I structure, CHES, coming from the crystallization buffer, is present in the centre of the active site, with the charged sulfonate group occupying a position equivalent to that of the halide ions in the structures of HheA, HheB and HheC (Watanabe et al., 2015) (de Jong et al., 2003) and equivalent to that of a water molecule in the structure of HheAD2 (de Jong et al., 2006). In the chain A, CHES molecule forms H bonds with the side chain of Ala-89 (3.3 Å), Met-144 (3.6 Å), Tyr-188 (3.5 Å) through its phenolic ring interactions with His-10 (3 Å) and His-9 (3.7 Å) of the His-tag used for the N-terminal His-tag can be seen. Amino group of CHES is H-bounded with the side chain of Asp-181 (2.8 Å). Finally it interacts through one oxygen of the sulfonate group with the side chain NH group of Glu-180 (2.6 Å), and Phe-39 (3.4 Å); through another oxygen with the side chain NH group of Asp-185 (2.8 Å) and with the third oxygen of the sulfonate group with side chains NH groups of Val-183 (2.9Å) and Phe-182 (2.9 Å) (**Fig.41-A**).

In HheD2 Crys_II a water molecule can be found at the center of the halide-binding site, occupying a position corresponding to that of the halide ions in the structures of two product complexes of HheC (de Jong et al., 2003) site, as in the case of HheAD2 structure (**Fig.41-B**). Such an *apo* structure is in agreement with the absence of halide salts

in the crystallization solution of Crys_II. The water molecule makes H-bonds with the side chain Tyr-150 (3.5 Å) and NH group of side chain Asp-181 (3.2 Å). It makes interactions with the side chain Ser-139 (4.2 Å) andanother water molecule (3.5Å).

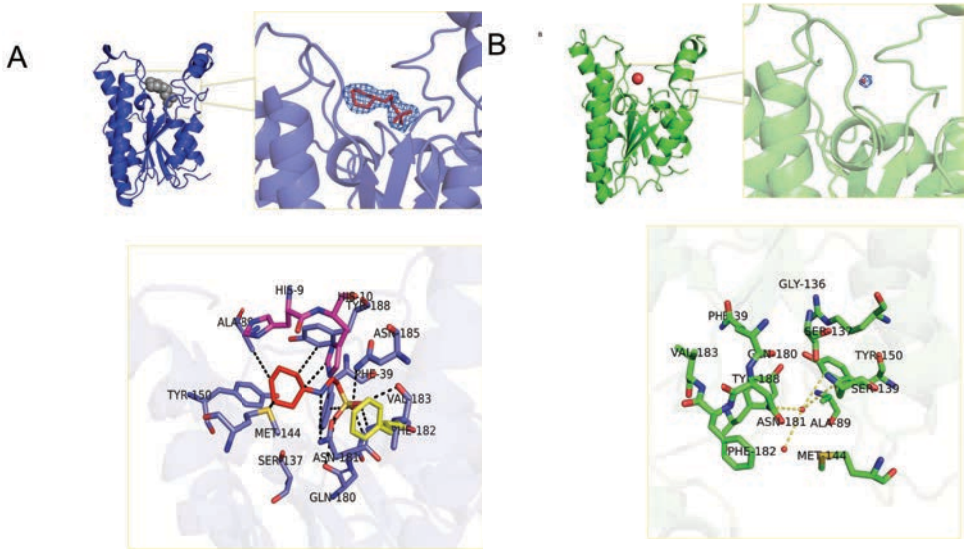


Figure 41: Electron density around the ligands present in the active sites of Crys_I and Crys_II structures and their coordination

Diagram illustrating the electron density map ($2F_0–F_c$ at 1σ) around CHES and water molecule found in Crys_I (**A**) and Crys_II (**B**) active sites.
On the right side a detailed overview of the protein active site, showing the CHES and water molecule coordination in the active site. In the case of CHES, many active site amino acids are involved in its coordination, and also an His (shown in magenta) coming from the neighbour monomer tag region that modifies the protein for purification on a nickel affinity column.

The HheD2 active site is very similar to the other HHDHs. The most notable difference regards the much more open structure of the substrate-binding pocket of the HheD2 with respect to the HheC. In fact, HheD2, lacking the C-terminal extension that contributes in HheC to the active site of the neighbouring subunit, has an active site much more exposed to the solvent (**Fig.42**). Another important contribution to the openness of the active site is the presence of a methionine (Met-44), whereas HheC has a tryptophan (Trp-139) at the equivalent position. Trp-139 is deemed to be responsible for the high enantioselectivity of HheC (Tang et al., 2003), precisely because the catalytic site is more restricted.

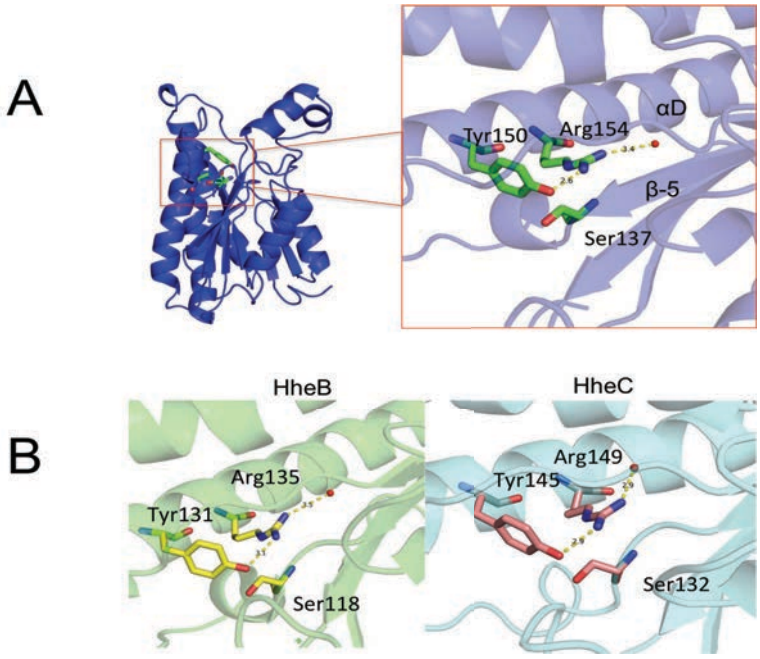


Figure 42: HheD2 catalytic triad, compared with HheB and HheC structures

(**A**) Cartoon view of HheD2 monomer, depicting the aminoacidic catalytic triad. On the right side, the three amino acids can be seen in zoom in stick representation. The catalytic water is shown as a sphere in red. The hydrogen bond between the Tyr-150 and the Arg-154 and between the Arg-154 and the W are also shown as dashed yellow lines.
(**B**) Catalytic triad from HheB (in green) and HheC (in cyan) are shown.

Similar to all the other H-Lyases structures, the active site contains the Ser137- Tyr150 -Arg154 catalytic triad. Tyr154 is hydrogen bonded to Arg-154 in the next turn of helix. The residues are located in adjacent turns of one of the two helices that form one of the dimer interfaces in the HheD2 tetramer, αD in the topology scheme in Fig.38. Tyr-150 is also bounded to a water molecule that is thought to relay the abstracted proton to the surface residue Asp-80. The third residue of the catalytic

triad, Ser-137, is located at the tip of β-strand 5 and it does not interact with any other residues of the triad.

The same interactions can be observed for the HheB and HheC, suggesting they share the same mechanism to generate the nucleophile species that substitutes the halogen of a bound haloalcohol substrate during catalysis. Further crystallisation studies of the HheD2 in complex with reaction intermediates are necessary to elucidate the catalytic mechanism in its details.

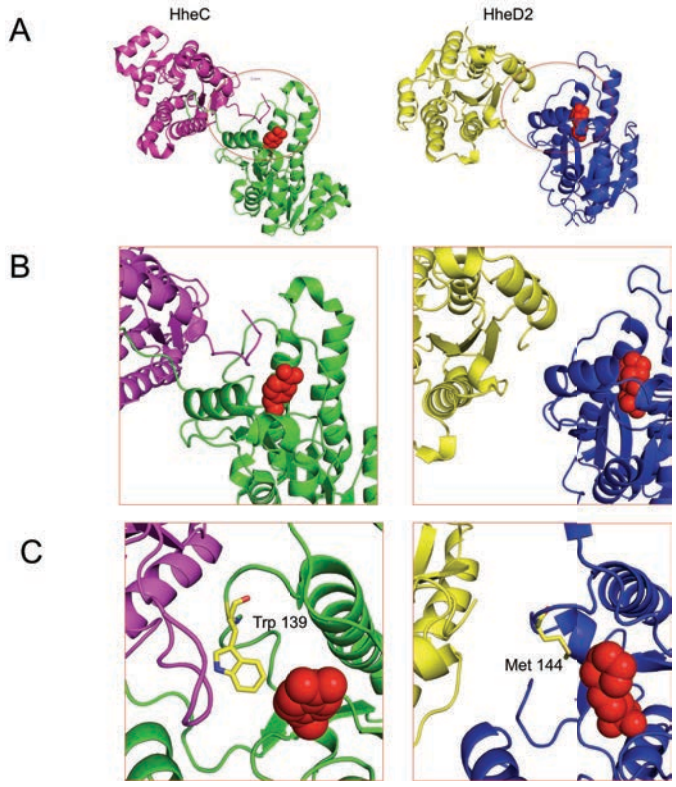


Figure 43: Comparison between HheD2 and HheC active sites

(**A**) In HheC the C-terminal extension protrudes into the active site of an opposite monomer within the tetrameric quaternary assembly. In HheD2 the C-terminal extension is absent and it contributes to the much open active site conformation (**B**) A zoom (**C**) The presence of a methionine residue along also contributes to the much open active site conformation.

In Crys_I structure His-9 and His-109, coming from the His-tag added to the protein for the purification strategy, are actually acting like the C-terminal extension in HheC, entering the active site of the next monomer.

The halide-binding site is formed by two loops extending from the βαβαβ Rossman fold motif. A loop starting from β-strand βF and a small loop between β-strand βA and helix αB form a spacious halide binding site.This loop is equivalent to a highly variable loop that is responsible for cofactor and substrate binding in members of evolutionarily related SDR family (Oppermann et al., 2003) (de Jong et al., 2006).

2.10 Catalytic mechanism of HheD2

The amino acids that constitute the catalytic triad play a central role in the dehalogenase reaction.

In HheC structure Try-145 abstracts a proton from the hydroxyl group of the substrate, acting as a catalytic base, and concomitantly the substrate oxygen performs a nucleophilic attack on the neighbouring halogen-substituted carbon atom, resulting in the ring closure and liberation of halide. Arg-149 does not interact with the substrate itself but lowers the pKa of the tyrosine-OH via hydrogen bonding in order to active it for proton abstraction (de Jong et al., 2006) (Janssen, 2007). Also in HheD2 the direct interaction of Arg-154 with Tyr150 (2.8 Å), in the otherwise hydrophobic active-site cavity likely polarizes the tyrosine hydroxyl group and lowers its pKa. Deprotonation of the tyrosine would allow this residue to act as the catalytic base that

accepts a proton from the hydroxyl group of a vicinal haloalcohol substrate to generate an intramolecular nucleophile, which concertedly substitutes the halogen. As demonstrated by mutational studies conducted on HheC, a substitution of Arg by Lys results in a loss of activity of 400-fold with respect to the *wild-type (van Hylckama Vlieg et al., 2001).* Impaired proton transfer between the tyrosine and the lysine could explain that.
The conserved Ser132 could be hydrogen bonded, thus ensuring proper positioning of the substrate as in the SDR family.

2.11 General considerations and future objectives

Halohydrin dehalogenases are remarkable enzymes with an exceptional potential for biocatalysis. Recently, a motif-based sequence database mining approach resulted in the identification of 37 novel halohydrin enzymes (Schallmey et al., 2014), and the expansion of the enzyme family by at least four novel enzyme family subtypes, D through G. As the majority of novel enzyme sequences significantly differs from those previously known, the novel HHDHs were expected to exhibit also novel biochemical and biocatalytic characteristics in comparison to the previously known HheA, HhAD2, HheB, HheC. In the group of Prof. Annett Schallmey from the Institute for Biochemistry, Biotechnology and Bioinformatics in Braunschweig (Germany), HHDH´s potential to convert the deoxyribose phosphate aldolase (DERA) products in chiral epoxide, which are valuable and essential building blocks for the synthesis of enathiomerically pure pharmaceuticals and agrochemicals (Choi et al., 2000), was tested.

HheD2 was the most promising candidate. The crystal structure of HheD2, solved by Arcimboldo_Shredder, revealed an overall organization very similar to the other HHDHs from different families, present in the PDB. Some remarkable differences were identified: the lack of additional secondary structure elements: HheD2, as HheAD2 and HheB, lacks a β-strand and a α-helix of HheC. Another significant difference in HheD2 is the absence of the C-terminal extension that donates a tryptophan residue (Trp-249) to the active site of an adjacent subunit in HheC. The halide binding of HheD2 is formed by two loops extending from βαβαβ Rossman fold motif. The first and largest one is equivalent to highly variable loop that is responsible for cofactor and substrate binding in members of the evolutionary related SDR family, whereas the second loop corresponds to the Gly-rich cofactor-binding motif of the SDR enzyme (de Jong et al., 2003) (Oppermann et al., 2003). In the absence of cofactor, it is unclear what the role of this loop may be, but as an enzyme involved in the degradation of a xenobiotic, it seems plausible that it has not yet fully evolved to an optimum. Moreover, similar to all the other H-Lyases structures, the active site contains the Ser137- Tyr150 -Arg154 catalytic triad. Analysing the active site architecture in comparison with the others known HHDHs, we can suppose a common catalytic mechanism, involving the catalytic triad. The present study provides insight in the structure of a new halohydrine dehalogenases D2 from Gammaproteobacterium. The structure will guide site-direct mutagenesis experiments on the wild-type protein designed to enhance the catalytic activity. In the future, the analysis of the protein and its variants in complex with reaction substrates will be carred on in order

to elucidate the catalytic mechanism and the role of all the species involved.

Experimental procedures

2.12 HheD2 expression

The HheD2 was expressed in the research group of Prof. Annett Schallmey, from the Institute for Biochemistry, Biotechnology and Bioinformatics in Braunschweig (Germany). The plasmid, which contains HheD2 gene, was introduced into BL21-DE3 pLys-S *E. coli* expression strain with pET28-TFAM. 2 µL (usually at 100 ng/µL) of the plasmid were deposited at the bottom of the microtube, onto which 50 µL BL21-DE3 pLys-S strain were added and incubated 30 min on ice. It followed a heat-shock step (1 min at 42º C) and 5 min of incubation on ice. Subsequently, 900 µL of LB medium were added to the transformation reaction and the cells were shacked at 200 rpm for 1h at 37º C. 100 µL of reaction were plated on an agar plate with kanamycin and incubated overnight (O/N) at 37º C. One colony of the transformation plate was inoculated into a pre-culture of 4 ml of LB and shaked at 250 rpm O/N at 37º C. The pre-culture was used to inoculate a 500 ml of LB large-scale culture, which was shacked at 200 rpm at 37º C, until it reaches an OD of 0.6-0.7. At this point the protein expression was induced with 1mM IPTG. After induction the culture are let grow for 4-5h at 37º and subsequently are harvested by centrifugation at 4000 rpm at 4º C for 25 min. The bacterial pellet can be stored at -80º C.

2.13 HheD2 purification

All the further steps were carried out at 4°C.

HheD2 cell pellet was re-suspended in 50 mL of lysis buffer A, containing 50 mM HEPES ph8, 500 mM NaCl, 25 mM imidazole, and cells were lysed using a cell disrupter (Constant Systems) at 1.35 Kbar.

Cellular debris was removed by centrifugation at 20,000 rpm for 30 min.

The clear supernatant was collected, centrifuged at 13,000 rpm for 20 minutes (Sigma 2 16-PK) and subsequently filtrated. The filtrated sample is then loaded and purified by affinity chromatography in an FPLC system (Amersham biosciences), using a Nickel Sepharose High Performance (GE Healthcare) containing HisTrap chelating stationary phase (Amersham Biosciences) with 5 mL bed volume and then by Size Exclusion Chromatography (SEC) in the same system, using a Superdex-200 gel filtration column (GE Healthcare).

2.13.1 Nickel Affinity

First a washing step of 5 column volumes with distilled water was performed. Then the column was equilibrated with lysis buffer A (5 column volumes at a flow rate of 3 mL/min. The filtrated supernatant ([6,8 mg/mL]) was applied to the column (divided in two runs) and subsequently washed with 25 mL of buffer B, containing 50 mM HEPES ph8, 500 mM NaCl, 100 mM imidazole at a flow rate of 1 mL min/min. Finally the protein was eluted with buffer C, containing 50 mM HEPES ph8, 500 mM NaCl and 500 mM imidazole, at a flow rate of 1 mL/min.

Two fractions of 5 ml, corresponding to the elution pick were collected for the run number 1 and two fractions of 3 mL for the run number 2.

The eluted fractions were analysed by SDS-PAGE using 15% polyacrylamide gels. The purest fractions were selected, pooled and concentrated up to 6 mg/mL in a total volume of 800 µL, using microfiltrators (Vivaspin 2 PES) with a cutoff of 5,000 Daltons. The

concentration measured by nanodrop (Nanodrop 2000 by Thermo Fisher Scientific). The absence of nucleic acids is assessed by the O.D. relation at 280 and 260 nm (OD^{260}/OD^{280}) of the peak fractions, which must be below 0.7.

2.13.2 Size exclusion chromatography (SEC)

The concentrated fractions were filtered (Millex-GP, 0.22 μm) and further purified on a Superdex-200 gel filtration column (GE Healthcare) equilibrated with buffer C. The protein was eluted using the same buffer at a flow rate of 1 mL/min. Peak fractions corresponding to a tetramer were pooled and concentrated up to 13 mg/ml using a 5,000 Daltons cutoff VIVASPIN 2. The protein was, finally, dialysed against HEPES 20 mM pH 7.5 and it was frozen at −20°C.

2.14 HheD2 crystallization

HheD2 protein (13 mg/ml) in a buffer HEPES (20 mM, pH 7.5) rendered bipyramidal crystals in lithium sulfate 0.2 M, CHES 0.1 M pH 9.5, and Na/K tartrate, using the sitting drop vapour-diffusion method at 293 K; and it rendered cubic crystals in sodium acetate 0.2M, Tris 0.1M pH 8,5, 30% PEG 4000. The mixture, consisted of equal volumes of protein with precipitant solution (2 μL), was equilibrated against 500 mL of the reservoir solution. The crystals, which appeared after three and seven days respectively, were soaked with a cryo-protectant solution (1.2 M sodium citrate and 20–25% glycerol), before flash cooling to 100 K.

The suitable crystallization conditions were found by extensive screening of chemical conditions employing automated dispensing methods. For this, 96 wells plates (Hampton Research) were dispensed by a nanodrop dispensing robot (Cartesian model by "Genomic solutions" or Phoenix by "Rigaku"). Nanodrops containing 100 nl of protein + 100 nl of crystallization buffer from the reservoir were prepared and monitored during time. The successful or promising crystallization conditions were then scaled up to µL volumes using 24 well plates (Hampton Research) also using the sitting drop set up, as explained.

2.15 X-ray diffraction, data processing and structure determination

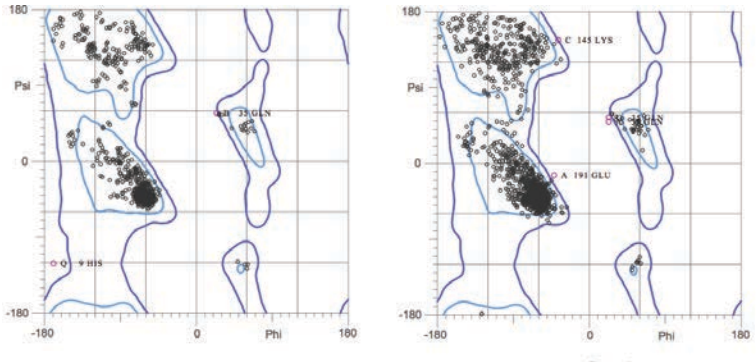
Two datasets of HheD2 with a total of 900 images spanning 180° were collected in frames of 0.2° rotation and with an exposure time of 0.1 min, using a PILATUS 6M detector and a fixed wavelength. The datasets were collected from the beamline XALOC at the ALBA synchrotron in Barcelona, Spain.

In both cases diffraction data were processed with XDS program. Intensities were scaled with SCALA and truncated to structure factors F with C-truncate, both from CCP4 suite.

For the dataset corresponding to the crystal type I the phase problem was solved using Arcimboldo Shredder.

For the dataset corresponding to the crystal type II the structure was solved by MR technique employing the structure solved for the crystal type I as searching model. Phaser as implemented in the Phenix suite

was used as MR program. The model was built manually with COOT and automatically refined with Phenix.refine and BUSTER.



Number of resiued in favoured region:
873 (96.36%) 1643 (95.41%)
Number of residue in allowed region:
31 (3.42%) 72 (4.18%)
Number of residues in outlier region:
2 (0.22%) 7 (0.41%)

Conclusions

CHAPTER 2: Structural solution of Halohydrin dehalogenases D2 (HheD2) from *Gammaproteobacterium.*

I. The full-length HheD2 of 25 kDa was overexpressed and purified. Through extensive screenings, the crystallization conditions were found and suitable protein crystals for X-ray diffraction were produced in two different buffers.

II. Two datasets were collected at 1.9 and 1.8 Å resolution at the ALBA synchrotron (Barcelona, Spain). The structure was solved using the fragment-based phasing method, Arcimboldo_Shredder.

III. The analysis of HheD2 reveals that it forms tetramers. One monomer is constituted by six β-strands that form a large parallel β-sheet structure, interacting through hydrophobic and electrostatic interactions of their side chains. A total of six alpha helices are distributed on both sides of the β-sheet, which resembles the characteristic Rossman fold, typical of the SDR related family.

IV. HheD2 shows an overall structure organization very similar to that of the other HHDHs available in the PDB, with some differences regarding the lack of some secondary structure elements present in the HheC (β-strand βC and an α-helix αC) and the absence of the C-terminal extension that links a tryptophan residue (Trp-249) to the active site of an adjacent subunit in HheC, which is responsible for the enanthioselectivity in this protein.

V. The halide-binding site is formed by two loops extending from the βαβαβ Rossmann fold motif. Similar to all the other H-Lyases structures, the active site contains the Ser-137, Tyr- 150 and Arg-154 catalytic triad. The similarity in the active site organization suggests that the catalytic mechanism could be shared with the other known HHDHs.

VI. The structure will guide site-direct mutagenesis experiments on the *wild-type* protein designed to enhance the catalytic activity.

VII. Further analysis of the protein and its variants in complex with reaction substrates are proposed in order to elucidate the catalytic mechanism and the role of all the species involved.

Appendix

A.1 Methods in protein production

A.1.1 Transformation with heat shock

Transformation of plasmid DNA into *E. coli* using the heat shock method is a basic technique of molecular biology. It consists of ininserting a foreign plasmid or ligation product into bacteria, using a calcium rich environment provided by calcium chloride to counteract the electrostatic repulsion between the plasmid DNA and bacterial cellular membrane. A sudden increase in temperature creates pores in the plasma membrane of the bacteria and allows for plasmid DNA to enter the bacterial cell. After a short incubation on ice, a mixture of chemically competent bacteria and DNA is placed at 42°C for 45 seconds (heat shock) and then placed back on ice. Super Optimal Broth Media (SOC) is added and the transformed cells are incubated at 37°C for 1 hour with agitation. To ensure isolating colonies irrespective of transformation efficiency, two quantities of transformed bacteria are plated. This traditional protocol can be used successfully to transform most commercially available competent bacteria.

A.1.2 Site directed mutagenesis with specific oligonucleotides

Site-directed mutagenesis studies can be extremely useful for elucidating the function of a gene or protein, or for creating variants of an enzyme with new and improved functions.

In this technique, a nucleotide sequence of interest is experimentally altered using synthetic oligonucleotides. The most commonly used approach is to use an oligonucleotide that is complementary to part of

a single-stranded DNA template, but containing an internal mismatch to direct the mutation.

In brief, point-mutations can be introduced to plasmids using primers (with the desired mutation) in a PCR protocol that amplifies the entire plasmid template. The parent template is removed using a methylation-dependent endonuclease (i.e. DpnI), and bacteria are transformed with the nuclease-resistant nicked plasmid (the PCR product). Plasmids are isolated from the resulting colonies, and screened for the desired modification. Finally, the positive clones are sequenced to confirm the desired modification and the absense of additional modifications.

A.2 Methods in protein purification and analysis

Isolation techniques and protein purification are essential requirements for the structural studies of proteins. The key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements and combine them in order to maximize yield and minimize the number of steps required, obtaining a suitable amount of pure protein (>90%).

A.2.1 Chromatography technique

Chromatography is a method for protein separation, based on differences in their structure and/or composition. In fact to be able to isolate a specific protein from a crude mixture the physical and/or chemical properties of the individual protein must be utilized.

In general, chromatography involves moving a mixture of the materials, dissolved in a fluid called mobile phase to be separated over a stationary phase. The molecules in the mixture will undergo different

interactions with the stationary support and, for this reason, they will travel at different speeds. The technique is based on the principle that molecules that have stronger affinity with the support will move more slowly through the support than those molecules with weaker interactions. In this way, different types of molecules can be separated from each other as they move over the support material.

In biotechnology, the most commonly used type of chromatography is liquid chromatography, which is a separation technique in which the mobile phase is a liquid. Present day liquid chromatography that generally utilizes very small packing particles and a relatively high pressure is referred to as high performance liquid chromatography (HPLC). In HPLC the sample is forced by a liquid at high pressure (the mobile phase) through a column that is packed with a stationary phase composed of irregularly or spherically shaped particles, a porous monolithic layer, or a porous membrane.

It is useful developing a purification strategy for the protein in question. Most purification protocols require more than one step to achieve the desired level of product purity. This includes any steps necessary to transfer the product from one technique into conditions suitable to perform the next technique. Each step in the process will cause some loss of product, therefore it is important to limit the number of steps in a purification procedure.

In this thesis work three different chromatography techniques have been used: nickel affinity, anion exchange and size exclusion.

A.2.1.1 Nickel affinity chromatography

Separation of a desired protein using affinity chromatography relies on the reversible interactions between the protein to be purified and the

affinity ligand coupled to chromatographic matrix. Most of the proteins have an inherent recognition site that can be used to select the appropriate affinity ligand. The binding between the protein of interest and the chosen ligand must be both specific and reversible. The process involves several steps. First, samples are supplied under conditions that favor maximum binding with the affinity ligand. After sample application, a washing step is performed to remove unbound substances, leaving the desired bound molecule still attached to the affinity support. To release and elute the bound molecules, a desorption step is usually performed either 1) specifically using a competitive ligand or 2) non-specifically by changing the media atmosphere (e.g. changing the ionic strength, pH or polarity) (Zachariou, 2008). As the elution is performed, the purified protein can be collected in a concentrated form.

Ni-affinity chromatography, in particular, uses the ability of histidine (His) to bind nickel. Six histidine amino acids at the end of a protein (either N- or C-terminus) constitute a 6X His tag. Nickel is bound to an agarose bead by chelation using nitrilotriacetic acid (NTA) beads. Several companies produce these beads as His Tagged proteins are currently some of the most used affinity tags.

A.2.1.2 Size exclusion chromatography (SEC)

This technique separates proteins based on size (hydrodynamic radius) and shape. Among the different subtypes of SEC, there is the gel filtration chromatography (GF) The support for GF is a resin made of beads which contain holes, called “pores”, of given sizes. Larger

molecules, which cannot penetrate the pores, travers the column around the beads and migrate through the spaces which separate the beads faster than the smaller molecules, which may penetrate the pores.

Generally, when the purity level of the target protein is not high enough, a polishing size exclusion chromatography is employed after the affinity purification step. SEC can be also useful for analytical purposes.

A.2.2 Protein analysis, SDS-PAGE electrophoresis

SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) separates proteins according to their molecular weight, based on their differential rates of migration through a sieving matrix (a gel) under the influence of an applied electrical field. Because molecules in an electric field move with a speed dependent on their charge, shape and molecular radius, electrophoresis has been extensively developed for molecular separations. In their native state, different proteins with the same molecular weight would migrate at different speeds in an electrical field depending on their charge and 3D shape. Then, in order to separate proteins in an electrical field based on their molecular weight only, it is necessary to eliminate effects derived of their structure by reducing the protein to a linear molecule, and masking the intrinsic net charge of the protein. Sodium dodecyl sulfate (SDS) is a detergent, that is present in the SDS-PAGE sample buffer, containing a reducing agent (normally DTT or B-ME to break down protein-protein disulphide bonds) and it is used to disrupt the tertiary

structure of the protein and to coat the protein with a uniform negative charge, which masks the intrinsic charges on the R-groups. At the end of separation, the different species are detected as bands at different position in the matrix. The most common analytical procedure for detection is staining. Proteins are usually stained with Comassie Brillant Blue.

A.3 Methods in protein crystallography

A.3.1.Protein crystallization

The aim of protein crystallization is to produce well-ordered mono protein crystals large enough to diffract X-ray beams. Like many other types of molecules, proteins can be prompted to form crystals when the solution in which they are dissolved becomes supersaturated. Under these conditions, individual protein molecules can pack in a ordered repeating array, held together by non-covalent interactions.

The crystallization process can be illustrated by a phase diagram that indicates which state (liquid, crystalline or amorphous solid precipitate) is stable under a variety of crystallization parameters. The **Fig.44** shows a typical crystallization phase diagram, consisting of three zones representing different degrees of supersaturation. Since there is an energy barrier, nucleation (the process of forming a nucleus) takes time. If the supersaturation is too small, the nucleation rate will be so slow that no crystals form in any reasonable amount of time. The corresponding area of the phase diagram is known as the “metastable zone.” In the “labile” or “crystallization” zone, the supersaturation is large enough that spontaneous nucleation is observable. If the supersaturation is too large, then disordered

structures, such as aggregates or precipitates, may form. The "precipitation zone" is unfavorable for crystal formation, because aggregates and precipitates form faster than the crystals. Even though the division in zones is qualitative, the different behaviors serve as guide when searching for the appropriate conditions to produce crystals (Mikol *et al*., 1989).

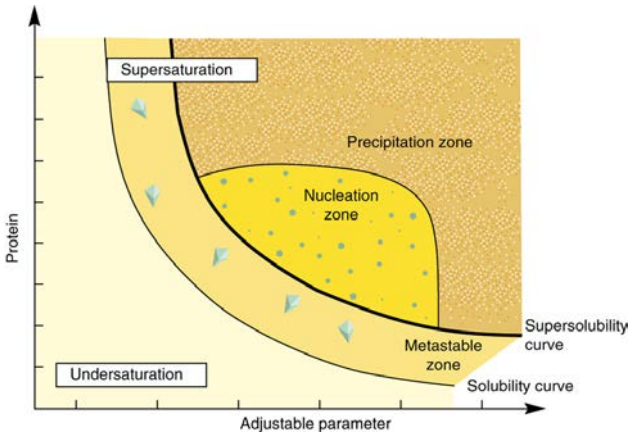


Figure 44: Crystallization phase diagram

The supersolubility curve is defined as the line separating conditions where spontaneous nucleation (or phase separation, precipitation) occurs from conditions where the crystallization solution remains clear if it is left undisturbed.
This figure was adapted from Chayen, N.E., Turning protein crystallization from an art into a science, Curr. Opin. Struct. Biol.,14, 577–583.

The first step to find the proper crystallization conditions is to set up screening trials. The protein solution will be mixed with a variety of buffers in order to find promising "hits". After that, a second screen with more defined range of conditions can be performed. There are

different methods to carry out crystallization experiments. In this thesis work the hanging and sitting drop vapor diffusion methods and seeding techniques have been employed.

A.3.1.1 Vapor diffusion

In this technique a small droplet of 2-10 μL of the protein is mixed with an equal volume of the crystallization buffer. The difference in concentration between the drop and the reservoir, that typically contains a higher reagent concentration than the drop, drives the system toward equilibrium by diffusion through the vapor phase. The protein crystals start to form when the drop and reservoir are at or close to equilibrium. In the sitting drop technique the protein drop is placed on the well in vapor equilibrium with the crystallization buffer. In the hanging drop technique the droplet is placed, together with the crystallization buffer on a siliconized glass cover slide inverted over the reservoir in vapor equilibrium with the reagent.

A.3.1.2 Seeding

The probability of spontaneous nucleation in nature increases with the degree of supersaturation. Nucleation can also be stimulated by seeding with crystals of the same protein grown under other conditions. A seed provides a template on which further molecules can assemble. Energetically, it is more favorable to add to an already existing crystal plane than it is to create a new nucleus. Given proper environment, time, and patience, the seed will enlarge into a crystal.

There are different ways of carrying out a seeding crystallization experiment: streak seeding, microseeding and macroseeding.
In this thesis work, the streakseeding technique has been performed.
The streakseeding consists in taking a thin support such as a whisker and touching a crystal to dislodge seeds. The seeds remain attached to the whisker and are transferred to a new protein-buffer drop as the whisker is drawn in a straight line across it. This produces a local increase in precipitant concentration along the streak line and it can stimulate nucleation in the absence of seeds.

A.3.1.3 Cryo-crystallography

Proteins are sensitive to X-ray radiation damage. The X-rays are ionizing radiation and their absorption creates radicals, which rapidly destroy any protein crystal. An efficient way to reduce radiation damage by slowing down the kinetics of the radical reactions is cryogenic cooling. This technique consists in fresh-cooling crystals to liquid nitrogen temperaturas (100K), either in cold nitrogen gas streams or directly into liquid nitrogen and is widely used as it strongly reduces radiation damage.

A.3.2. X-ray diffraction technique

A.3.2.1. X-ray diffraction principles

The most common experimental method of obtaining a detailed model of a large molecule, reaching the resolution of individual atoms, is to interpret the diffraction pattern of X-rays scattered from many copies of the same molecule distributed in an orderd array called a crystal. This method is named single crystal X-ray crystallography. X-rays are electromagnetic radiation with typical photon energies in the range of 100 eV -100 keV. Because the wavelength of X-rays is comparable to the size of atoms, they are suitable for probing the structural arrangement of atoms in the crystal. A crystal of a protein contains many ordered molecules in identical orientations, so each molecule diffracts identically, and the diffracted beams for all molecules may constructively interfere to produce strong, detectable X-ray beams. The periodic arrangiament of a crystal, the motif, is placed in a symmetrical manner on a 3D network, called lattice. The smallest repeating volume of the crystal is referred as unit cell (u.c.) whose translations in the 3d space generate the lattice network. The u.c. is described by three lattice constants a, b, c and by three angles α, β, γ. When a beam of X-rays hits a crystal, the electrons within the crystal diffract the incident X-rays into specific directions. The ordered atoms in the crystal, which are disposed in planes, when diffracted, generate an elastic scattering. A regular array of scatterers produces a regular array of spherical waves. Only waves that add constructively will contribute to the diffraction pattern. In order for the waves to interfere constructively, the differences in the travel path must be equal to integer multiples of the wavelenght. When this constructive interference occurs, a

diffracted beam of x-rays will leave the crystal at an angle equal to that of the incident beam (θ). By knowing the phase angle and intensity of the diffracted X-rays, it is possible to generate a three-dimensional representation of the electron density within the crystal. Diffraction spots are formed according to the Bragg's law equation:

$$\mathbf{2d \times \sin\theta = n\lambda}$$

In this equation, *d* is the spacing between diffracting planes, θ is the incident angle, *n* is any integer, and λ is the wavelength of the beam.

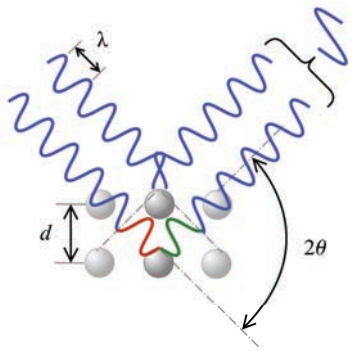


Figure 45: Bragg's law

Diagram of Bragg's Law angle of deviation 2θ

The scattered X-ray from the crystal allows the measurement of a large number of Bragg reflections on a bidimensional detector in a single exposure. A crystal lattice is 3D, only a fraction of the crystal lattice points are in diffracting position at any given orientation of the crystal. Therefore, the crystal is also rotated on single or multiple axis through an angle of 0.1 to 2° during the exposure to bring more reflections to diffracting position. The crystal lattice presents a rotational symmetry accrding the Laue group of the space group in which the molecules has crystallized. Thanks to this property is not mandatory, although is

recommended, probing the whole space of orientations while collecting data. The process of combining and merging all collected images and reflections in a single data set is called *integration*.

The discrete planes, which give rise to reflections, are called lattice planes, and have an orientation relative to the lattice defined by Miller indices, *hkl*.

These indexes define each point that constitutes an *hkl* reflection with a specific intensity I_{hkl}. The mathematic Ewald's sphere model shows the correlation between the crystal diffraction and its reciprocal space lattice construction. This sphere is a representation in reciprocal space of all the possible points where planes (reflections) could satisfy the Bragg's equation.

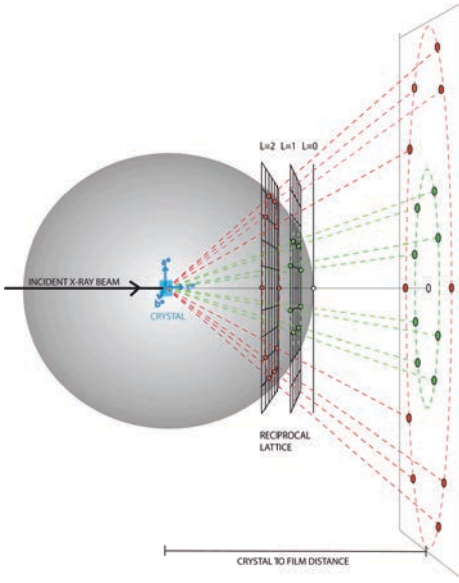


Figure 46: Ewald sphere

The diffracted ray can go in any direction in three dimensions, so the vector representing it can have its tip anywhere at the surface of a sphere with a radius of 1/λ. Such a sphere, called the Ewald sphere, is shown in the figure on the left.

In order to define the Miller indices of each reflection, an indexing procedure is required to identify the unit cell dimension and the crystal symmetry constraints. These, along with the intensity statistics will

allow to infer its possible space group. After indexing, the spot area relative to the same reflection is integrated from all images giving the final intensity value (I_{hkl}). It follows the scaling of the intensity by using a scale factor. The scale factor is required to normalize factors, by accounting for experimental conditions, such as beam intensity decay or crystal centering. Data integration and scaling will also require an error model to account for inconsistencies caused by radiation damage (determined by the X-ray on the protein), the instrument, the geometry of the experiment or the crystal itself. Once the scaled intensities (I) are obtained, they are truncated to amplitudes (F) by applying the premise of $F^2=I$.

Two separate pieces of information can be found in the reflections of the diffraction images. The first comes from the geometrical arrangement of the reflections, which gives all the information about the crystal lattice and the symmetry of the crystal. The second comes from the intensity of the reflection that gives part of the information about the content of the lattice. Unfortunately the second information lacks the phase associated to the diffracted X-ray beams, leading to the well-known crystallographic phase problem.

In X-ray crystallography the structure factor *F(hkl)* of any reflection hkl is the value that express both the amplitude, module |F(h,k,l)|, and the phase, Φ(hkl), of that reflection. In order to calculate the electron density map both parameters must be known.

The Fourier transform is the mathematical operation that converts discrete structure factors into a continuous 3D distribution of electron density in the real space.

To converts discrete structure factors into a continuous 3D distribution of electron density in the real space the Fourier transform convolution is applied following:

electron density function:

$$\boldsymbol{\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F(h,k,l)|exp[-2\pi i * ((hx + ky + lz) - i\Phi(hkl))]}$$

Here ρ is the electron density; x, y, and z are the cartesian coordinates in the real space, V is the volume of the unit cell, Fhkl the structure factor for a given set of Miller indices, is an imaginary number and Φ the phase.

For solving the phase problem and obtaining initial phase estimates a number of different approaches can be used:

for example Ab initio phasing or direct methods – This is usually the method of choice for small molecules (<1000 non-hydrogen atoms), and has been used successfully to solve the phase problems for small proteins. If the resolution of the data is better than 1.0 Å, direct methods can be used to obtain phase information, by exploiting known phase relationships between certain groups of reflections (Hauptman *et al*., 1997; Usón *et al.*, 1999).

If a homolog structure is known, it can be used as a search model in molecular replacement to determine the orientation and position of the molecules within the unit cell.

Molecular Replacement method was used for solving of $SHMT_{Sth}$ structure in this work.

A.3.2.2 Molecular replacement (MR)

Molecular Replacement is a technique that aims to combine the phases from a homologous protein with the reflections of the unknown structure. As the homolog could have been solved in a different unit cell and space group it is required to find the correct orientatation (rotation and translation) of this model in the target crystal.

For a successful application of MR the probe should have a sequence identity with the target above 25-35%. The MR task involves positioning the probe within the unit cell of the target crystal in such a way that the theoretical diffraction that would result from this model corresponds the experimental one. This idea underlies a variety sophisticated mathematical functions, which are used in practice (ref rossman 72).

We have a template molecule A and the unknown molecule A' homolog to A (**Fig.41**). The position of A' is different from A. To superimpose the molecule A with A' we have, firstly, to apply the rotation R, and then the translation T. Therefore, the main aim of the MR method is to find these two operators, or, in other terms, to solve the Rotation and the Translation functions.

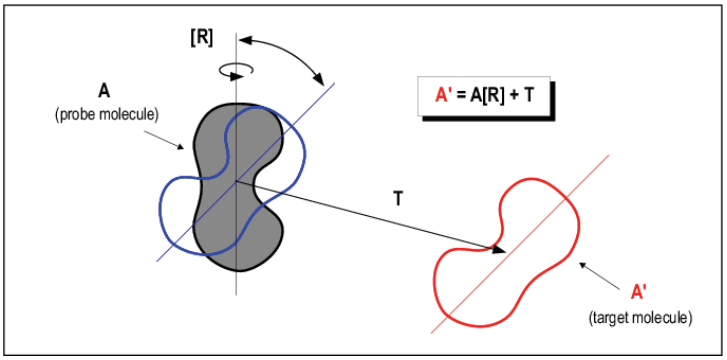


Figure 47: Molecular replacement method visually explained.

A representation of the main principle of molecular replacement method. **A** is the probe molecule, and **A'** is the target molecule. The operation, consisting of rotation and translation, superimposes the probe onto the target molecule as shown. [R] is the rotation matrix, and T is the translation vector.

Rotation and translational functions can be performed by following two different approaches: the traditional Patterson method and the probability approach (Evans and McCoy, 2008). In the traditional Patterson method the rotation searches are based on the Patterson function, scoring the overlap between target and model Patterson maps in the region close to the origin, where the function is dominated by intramolecular self-vectors that are independent from translation. After rotation, in the translation search, the now correctly oriented known model can be correctly positioned by translating it to the correct co-ordinates within the asymmetric unit. This is accomplished by moving the model, calculating a new Patterson map, and comparing it to the target Patterson map.

The modern probability approach uses the 'Maximum-likelihood' statistical methods (Read R. J. et al., 2007). This method take into account the errors in the experimental data as well as the similarity

degree between the model and the target protein in order to be able to estimate the error probability that is due to a low similarity of the model and not to a wrong solution. In a maximum-likelihood rotation search, the model is rotated sequentially on an angular grid and the orientation that predicts the data with highest probability is selected. As for the rotation search, the translation search for any given orientation consists in placing the model sequentially in grid points throughout the translationally unique volume of unit cell. Intensity based formulae (Read, 1990) give the probability for each putative rotation and translation, from which the most likely are selected as the solution.

If the MR process gives one solution that clearly stands out in scores from the next best result, it is likely to be correct. The principal test for a correct and useful solution is that the maps calculated from the solution model should show new and plausible information that was not present in the model. This might be side chains or loops that were different in the model and are present in the target protein. Moreover high values of ZSCORE (above 8) indicate a clear solution.

A.3.2.3 Refinement and validation methods

After the initial phasing and building, the model of a protein is generally far from perfect. To improve the phases and also the interpretation of the electron density map, refinement methods are a very important step in the interpretation of the diffraction data. Refining is achieved through statistical adjustment of the atomic coordinates to fit the diffraction data better. As a measure of the fitting the "work R-factor" is used, that measures how far the calculated

amplitudes differ from the observed amplitudes. The crystallographic R-factor is defined by the equation:

$$R = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} F_{hkl}^{obs}}$$

In addition to the R-factor, the so called free R-factor can be used as a cross-validating indicator to monitor the overall progress and to avoid fitting noise. In fact, the R-factor can get trapped in local minima giving the false impression of having a good model. The free R-factor is calculated from a set of the reflections (typically 5%) that are excluded from refinement and it therefore gives an independent measure of the refinement progress. The drop of the R-factors during the refinement is more important than the actual value, but generally final values lie between 15% and 25% with the free R-factor being slightly higher than the work R-factor

A.3.3 Arcimboldo-Schredder: a method for structure solution using distant homologs as models.

MR can solve macromolecular structures when suitable search models are available. Models from distant homologs may deviate too much from the target structure to succeed. Successful ways to make the most of such templates usually rely on the degree of sequence conservation to select and improve search models. ARCIMBOLDO_SHREDDER (Millán et al., 2015b) uses fragments derived from distant homologs in a brute force approach driven by the experimental data, instead of by sequence similarity.

In fact, in ARCIMBOLDO_SHREDDER, fragments are derived from a distant homolog template, and their performance is jointly evaluated

in a process driven by the experimental data, provided a resolution of at least 2.5 Å is available. In its original implementation, template trimming relied on a rotation function based scoring, the SHRED-LLG. The whole template was initially used to find the maxima of the rotation function. The list of peaks in the rotation function was clustered within a given tolerance. For each of these clusters, the template was systematically shredded (omitting from the polypeptide chain continuous stretches in a range of sizes) and fragments were scored against each unique solution of the rotation function. Then, results were combined into a score per residue and the template was trimmed accordingly. The sequential shredding and its derived model trimming can improve models where the high average deviation to the target is due to dissimilar or flexible regions reducing the signal from a core of low root mean square deviation (r.m.s.d.).

The original implementation of this idea, which aimed to eliminate all the most incorrect regions in the starting model, has been further extended (Millán et al., in preparation). Small, compact models, of equal size, are derived from the initial template by a distance-based algorithm. Such partial models, making as little as 10% of the main chain atoms in the asymmetric unit need to be very accurate (around 0.5Å r.m.s.d.) for their correct placement and extension into the full structure at 2 Å resolution. In order to increase the radius of convergence of this approach, additional degrees of freedom are given to the models, which are decomposed and subject to refinement against the intensity based likelihood rotation function (Read and McCoy, 2016) and again after they have been placed in the unit cell. This refinement is accomplished in PHASER with the gyre and gimble modes (McCoy et al., 2017).

A.4 Methods in protein biophysical characterization

A.4.1 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is a technique in physics that can be used to determine the size distribution profile of small particles in suspension or polymers in solution.

DLS has become a very useful tool in protein crystallization, because it gives information about size distribution, stability and aggregation state of macromolecules in solution. DLS is also used to understand how experimental variables influence aggregation.

DLS is based on the extraction of spectral information derived from time-dependent fluctuations of the light scattered from dissolved macromolecules in solution. In fact, when a suspension of particles is hit by a monochromatic coherent beam of light, generated scattered light waves spread out in all directions. Scattered waves interference in the far field region generates a net scattered light intensity Is(t)Is(t). Due to the random motion (Brownian motion) of the suspended particles.

The suspended particles of the colloidal dispersion under investigation undergo Brownian motion. This motion results in fluctuations of the distances between the particles and hence also in fluctuations of the phase relations of the scattered light. Additionally, the number of particles within the scattering volume may vary in time.

Due to the fact that large molecules or particles move slower than small molecules a defined correlation function can be defined and from it the correlation coefficient (D) of the molecules can be

calculated. Finally the hydrodynamic radius (Rh) of the particles and molecules can be calculated as follows:

$$D = kT / 6$$

In the case of protein crystallography the polydispersity of the sample and the presence of protein aggregates are actually more interesting parameters than the absolute hydrodynamic radius, because these parameters can interfere with the crystallization process. The polydispersity is the relative standard deviation of a sample. A sample can be considered polydisperse, if its corresponding polydispersity is more than 30% and not polydisperse if this value is between 20% and 30%.

Bibliography

ANASTAS, P. & EGHBALI, N. 2010. Green chemistry: principles and practice. *Chem Soc Rev,* 39**,** 301-12.

ANDERSON, D. D., WOELLER, C. F., CHIANG, E. P., SHANE, B. & STOVER, P. J. 2012. Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. *J Biol Chem,* 287**,** 7051-62.

ARJAN, J., VAN DEN WIJNGAARD, D. B. & JANSSEN, B. W. 1989. Degradation of Epichlorohydrin and Halohydrins by Bacterial Cultures Isolated from Freshwater Sediment. *Microbiology,* 135**,** 2199-2208,.

BARTLETT, G. J., PORTER, C. T., BORKAKOTI, N. & THORNTON, J. M. 2002. Analysis of catalytic residues in enzyme active sites. *J Mol Biol,* 324**,** 105-21.

BHAVANI, B. S., RAJARAM, V., BISHT, S., KAUL, P., PRAKASH, V., MURTHY, M. R., APPAJI RAO, N. & SAVITHRI, H. S. 2008. Importance of tyrosine residues of Bacillus stearothermophilus serine hydroxymethyltransferase in cofactor binding and L-allo-Thr cleavage. *FEBS J,* 275**,** 4606-19.

BHAVANI, S., TRIVEDI, V., JALA, V. R., SUBRAMANYA, H. S., KAUL, P., PRAKASH, V., APPAJI RAO, N. & SAVITHRI, H. S. 2005. Role of Lys-226 in the catalytic mechanism of Bacillus stearothermophilus serine hydroxymethyltransferase--crystal structure and kinetic studies. *Biochemistry,* 44**,** 6929-37.

BLAKE, C. C., KOENIG, D. F., MAIR, G. A., NORTH, A. C., PHILLIPS, D. C. & SARMA, V. R. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature,* 206**,** 757-61.

BORNSCHEUER, U. T. & HESSELER, M. 2010. Enzymatic removal of 3-monochloro-1,2-propanediol (3-MCPD) and its esters from oils Authors. *European Jornal of Lipid Science and Technology*.

BORNSCHEUER, U. T., HUISMAN, G. W., KAZLAUSKAS, R. J., LUTZ, S., MOORE, J. C. & ROBINS, K. 2012. Engineering the third wave of biocatalysis. *Nature,* 485**,** 185-94.

BROVETTO, M., GAMENARA, D., MENDEZ, P. S. & SEOANE, G. A. 2011. C-C bond-forming lyases in organic synthesis. *Chem Rev,* 111**,** 4346-403.

BRUNGER, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* 355**,** 472-5.

CASTRO, C. E. & BARTNICKI, E. W. 1968. Biodehalogenation. Epoxiation of halohydrins, epoxide opening, and transhalogenation by a Flavobacterium sp. *Biochemistry,* 7**,** 3213-8.

CHAVEZ, A., FERNANDEZ, M., LERAYER, A., MIERAU, I., KLEEREBEZEM, M. & HUGENHOLTZ, J. 2002. Metabolic engineering of acetaldehyde production by Steptococcus thermophilus. *Appl Environ Micro***,** 68-5656.

CHITNUMSUB, P., JARUWAT, A., RIANGRUNGROJ, P., ITTARAT, W., NOYTANOM, K., OONANANT, W., VANICHTHANANKUL, J., CHUANKHAYAN, P., MAENPUEN, S., CHEN, C. J., CHAIYEN, P., YUTHAVONG, Y. & LEARTSAKULPANICH, U. 2014. Structures of Plasmodium vivax serine hydroxymethyltransferase: implications for ligand-binding specificity and functional control. *Acta Crystallogr D Biol Crystallogr,* 70**,** 3177-86.

CHOI, W. J., CHOI, C. Y., DE BONT, J. A. & WEIJERS, C. A. 2000. Continuous production of enantiopure 1,2-epoxyhexane by yeast epoxide hydrolase in a two-phase membrane bioreactor. *Appl Microbiol Biotechnol,* 54**,** 641-6.

CHRISTEN, P. & MEHTA, P. K. 2001. From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Chem Rec,* 1**,** 436-47.

CLAPES, P. 2015. Biocatalysis in Organic Synthesis.

CLAPES, P. 2016. Recent Advances in Enzyme-Catalyzed Aldol Addition Reactions. In Green Biocatalysis.

CONTESTABILE, R., ANGELACCIO, S., BOSSA, F., WRIGHT, H. T., SCARSDALE, N., KAZANINA, G. & SCHIRCH, V. 2000. Role of tyrosine 65 in the mechanism of serine hydroxymethyltransferase. *Biochemistry,* 39**,** 7492-500.

CORNISH-BOWDEN, A. 2014. Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspect. Sci.,* 1**,** 74-87.

DAVIS, I. W., LEAVER-FAY, A., CHEN, V. B., BLOCK, J. N., KAPRAL, G. J., WANG, X., MURRAY, L. W., ARENDALL,

W. B., 3RD, SNOEYINK, J., RICHARDSON, J. S. & RICHARDSON, D. C. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res,* 35**,** W375-83.

DE JONG, R. M., KALK, K. H., TANG, L., JANSSEN, D. B. & DIJKSTRA, B. W. 2006. The X-ray structure of the haloalcohol dehalogenase HheA from Arthrobacter sp. strain AD2: insight into enantioselectivity and halide binding in the haloalcohol dehalogenase family. *J Bacteriol,* 188**,** 4051-6.

DE JONG, R. M., TIESINGA, J. J., ROZEBOOM, H. J., KALK, K. H., TANG, L., JANSSEN, D. B. & DIJKSTRA, B. W. 2003. Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. *EMBO J,* 22**,** 4933-44.

DI SALVO, M. L., BUDISA, N. & CONTESTABILE, R. 2013a. PLP-dependent Enzymes: a Powerful Tool or Metabolic Synthesis of Non-canonical Amino Acids

DI SALVO, M. L., CONTESTABILE, R., PAIARDINI, A. & MARAS, B. 2013b. Glycine consumption and mitochondrial serine hydroxymethyltransferase in cancer cells: the heme connection. *Med Hypotheses,* 80**,** 633-6.

DOLETHA, M. E., SZEBENYI , XIAOWEN LIU , IRINA A. KRIKSUNOV , PATRICK J. STOVER & THIEL, D. J. 2000. Structure of a Murine Cytoplasmic Serine Hydroxymethyltransferase Quinonoid Ternary Complex: Evidence for Asymmetric Obligate Dimers. *Biochemistry,* 39 (44)**,** 13313–13323.

DUNATHAN, H. C. 1966. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc Natl Acad Sci U S A,* 55**,** 712-6.

ELENKOV, M. M., HAUER, B. & JANSSEN, D. B. 2006. Enantioselective Ring Opening of Epoxides with Cyanide Catalyzed by Halohydrin Dehalogenases: A New Approach to Non-Racemic β-Hydroxy Nitriles. *ChemInform.*

ELIOT, A. C. & KIRSCH, J. F. 2004. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem,* 73**,** 383-415.

EMANUELSSON, M. A., OSUNA, M. B., SIPMA, J. & CASTRO, P. M. 2008. Treatment of halogenated organic compounds and monitoring of microbial dynamics in up-flow fixed bed reactors

under sequentially alternating pollutant scenarios. *Biotechnol Bioeng,* 99**,** 800-10.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2126-32.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

EVANS, P. & MCCOY, A. 2008. An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr.,* 64**,** 1-10.

FENG, E., YE, D., LI, J., ZHANG, D., WANG, J., ZHAO, F., HILGENFELD, R., ZHENG, M., JIANG, H. & LIU, H. 2012. Recent advances in neuraminidase inhibitor development as anti-influenza drugs. *ChemMedChem,* 7**,** 1527-36.

FERRER, J., NEGNYA, S., CORTES ROBLES, G. & LE LANNA, J. M. 2012. Eco-innovative design method for process engineering. *Computers & Chemical Engineering,* 45**,** 137-151.

FERSHT, A. 1999. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. *New York: W. H. Freeman.*

FISCHER, J. D., HOLLIDAY, G. L., RAHMAN, S. A. & THORNTON, J. M. 2010. The structures and physicochemical properties of organic cofactors in biocatalysis. *J Mol Biol,* 403**,** 803-24.

FRAZZETTO, G. 2003. White biotechnology. *EMBO Rep,* 4**,** 835-7.

FU, T. F., RIFE, J. P. & SCHIRCH, V. 2001. The role of serine hydroxymethyltransferase isozymes in one-carbon metabolism in MCF-7 cells as determined by (13)C NMR. *Arch Biochem Biophys,* 393**,** 42-50.

GALLAGHER, D. T., GILLILAND, G. L., XIAO, G., ZONDLO, J., FISHER, K. E., CHINCHILLA, D. & EISENSTEIN, E. 1998. Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. *Structure,* 6**,** 465-75.

GARROW, T. A., BRENNER, A. A., WHITEHEAD, V. M., CHEN, X. N., DUNCAN, R. G., KORENBERG, J. R. & SHANE, B. 1993. Cloning of human cDNAs encoding mitochondrial and cytosolic serine hydroxymethyltransferases and chromosomal localization. *J Biol Chem,* 268**,** 11910-6.

GIARDINA, G., BRUNOTTI, P., FIASCARELLI, A., CICALINI, A., COSTA, M. G., BUCKLE, A. M., DI SALVO, M. L., GIORGI, A., MARANI, M., PAONE, A., RINALDO, S., PAIARDINI, A., CONTESTABILE, R. & CUTRUZZOLA, F. 2015. How

pyridoxal 5'-phosphate differentially regulates human cytosolic and mitochondrial serine hydroxymethyltransferase oligomeric state. *FEBS J,* 282**,** 1225-41.

GRISHIN, N. V., PHILLIPS, M. A. & GOLDSMITH, E. J. 1995. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci,* 4**,** 1291-304.

GUO, C., CHEN, Y., ZHENG, Y., ZHANG, W., TAO, Y., FENG, J. & TANG, L. 2015. Exploring the enantioselective mechanism of halohydrin dehalogenase from Agrobacterium radiobacter AD1 by iterative saturation mutagenesis. *Appl Environ Microbiol,* 81**,** 2919-26.

GUTIERREZ, M. L., GARRABOU, X., AGOSTA, E., SERVI, S., PARELLA, T., JOGLAR, J. & CLAPES, P. 2008. Serine hydroxymethyl transferase from Streptococcus thermophilus and L-threonine aldolase from Escherichia coli as stereocomplementary biocatalysts for the synthesis of beta-hydroxy-alpha,omega-diamino acid derivatives. *Chemistry,* 14**,** 4647-56.

GUTTERIDGE, A. & THORNTON, J. M. 2005. Understanding nature's catalytic toolkit. *Trends Biochem Sci,* 30**,** 622-9.

HAMLET, C. G., SADD, P. A., CREWS, C., VELISEK, J. & BAXTER, D. E. 2002. Occurrence of 3-chloro-propane-1,2-diol (3-MCPD) and related compounds in foods: a review. *Food Addit Contam,* 19**,** 619-31.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HE, X. M. & LIU, H. W. 2002. Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annu Rev Biochem,* 71**,** 701-54.

HEDSTROM, L. 2001. Enzyme Specificity and Selectivity.

HERNANDEZ, K., ZELEN, I., PETRILLO, G., USÓN, I., WANDTKE, C., BUJONS, J., JOGLAR, J., PARELLA, T. & CLAPÉS, P. 2015a. IEngineered L-Serine Hydroxymethyltransferase from Streptococcus thermophilus for the Synthesis of α,α-Dialkyl-α-Amino Acids. *Angew Chem Int Ed Engl*.

HERNANDEZ, K., ZELEN, I., PETRILLO, G., USON, I., WANDTKE, C. M., BUJONS, J., JOGLAR, J., PARELLA, T. & CLAPES, P. 2015b. Engineered L-serine hydroxymethyltransferase from Streptococcus thermophilus for the synthesis of alpha,alpha-dialkyl-alpha-amino acids. *Angew Chem Int Ed Engl,* 54**,** 3013-7.

HEYL, D. L., HARRIS, E. & FOLKERS, K. 1951. Phosphates of vitamin B6 group I. The structure of Co-decarboxylase. *J. American Chemical Society,* 73 3430-3433.

HIGGINS, T. P., HOPE, S. J., EFFENDI, A. J., DAWSON, S. & DANCER, B. N. 2005. Biochemical and molecular characterisation of the 2,3-dichloro-1-propanol dehalogenase and stereospecific haloalkanoic dehalogenases from a versatile Agrobacterium sp. *Biodegradation,* 16**,** 485-92.

HIRACUMA, S., VASSIL, P., TETSUYA, K. & CHI-HUEY, H. 1995. Design of oligosaccharide mimetics with higher stability and simpler structures that parent oligosaccharides. *Riken*.

HOLLIDAY G. L., J.M. THORNTON, A. MARQUET, A.G. SMITH, F. RÉBEILLÉ, R. MENDEL, SCHUBERT H.L., LAWRENCE A.D. & M., W. 2007. Evolution of enzymes and pathways for the biosynthesis of cofactors. 24**,** 972-87.

HYDE, C. C., AHMED, S. A., PADLAN, E. A., MILES, E. W. & DAVIES, D. R. 1988. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium. *J Biol Chem,* 263**,** 17857-71.

JAGER, J., MOSER, M., SAUDER, U. & JANSONIUS, J. N. 1994. Crystal structure of Escherichia coli aspartate aminotransferase in two conformations. Comparison of un unliganded open and two liganded closed forms. *J Mol Biol,* 239**,** 285-305.

JAIN, M., NILSSON, R., SHARMA, S., MADHUSUDHAN, N., KITAMI, T., SOUZA, A. L., KAFRI, R., KIRSCHNER, M. W., CLISH, C. B. & MOOTHA, V. K. 2012. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science,* 336**,** 1040-4.

JANSSEN, D. B. 2007. Biocatalysis by dehalogenating enzymes. *Adv Appl Microbiol,* 61**,** 233-52.

JOHANNES, T., SIMURDIAK, M. R. & ZHAO, H. 2005. Biocatalysis. *Encyclopedia of Chemical Processing.*

JOHN, R. A. 1995. Pyridoxal phosphate-dependent enzymes. *Biochim Biophys Acta,* 1248**,** 81-96.

JORNVALL, H., PERSSON, B., KROOK, M., ATRIAN, S., GONZALEZ-DUARTE, R., JEFFERY, J. & GHOSH, D. 1995. Short-chain dehydrogenases/reductases (SDR). *Biochemistry,* 34**,** 6003-13.

KABSCH, W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

KAO, F., CHASIN, L. & PUCK, T. T. 1969. Genetics of somatic mammalian cells. X. Complementation analysis of glycine-requiring mutants. *Proc Natl Acad Sci U S A,* 64**,** 1284-91.

KAVANAGH, K. L., JORNVALL, H., PERSSON, B. & OPPERMANN, U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci,* 65**,** 3895-906.

KERN, A. D., OLIVEIRA, M. A., COFFINO, P. & HACKERT, M. L. 1999. Structure of mammalian ornithine decarboxylase at 1.6 A resolution: stereochemical implications of PLP-dependent amino acid decarboxylases. *Structure,* 7**,** 567-81.

KLEYWEGT, G. J. & BRUNGER, A. T. 1996. Checking your imagination: applications of the free R value. *Structure,* 4**,** 897-904.

KLEYWEGT, G. J. & JONES, T. A. 1996. Phi/psi-chology: Ramachandran revisited. *Structure,* 4**,** 1395-400.

KOOPMEINERS, J., HALMSCHLAG, B., SCHALLMEY, M. & SCHALLMEY, A. 2016. Biochemical and biocatalytic characterization of 17 novel halohydrin dehalogenases. *Appl Microbiol Biotechnol,* 100**,** 7517-27.

KOSHLAND, D. E., JR., RAY, W. J., JR. & ERWIN, M. J. 1958. Protein structure and enzyme action. *Fed Proc,* 17**,** 1145-50.

KRISSINEL, E. 2015. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res,* 43**,** W314-9.

LIU, J. Q., DAIRI, T., ITOH, N., KATAOKA, M., SHIMIZU, S. & YAMADA, H. 2000. Diversity of microbial threonine aldolases and their application. *J. Mol. Catal. B: Enzym,* 10**,** 107-115.

MACFARLANE, A. J., ANDERSON, D. D., FLODBY, P., PERRY, C. A., ALLEN, R. H., STABLER, S. P. & STOVER, P. J. 2011. Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. *J Biol Chem,* 286**,** 44015-22.

MATTHEWS, B. W. 1968. Solvent content of protein crystals. *J Mol Biol,* 33**,** 491-7.

MATTHEWS, R. G. & DRUMMOND, J. T. 1990. Providing one-carbon units for biological methylations: mechanistic studies on serine

hydroxymethyltransferase, methylenetetrahydrofolate reductase, and methyltetrahydrofolate-homocysteine methyltransferase. *Chem. Rev.*, 90 (7)

, 1275–1290.

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.

MCCOY, A. J., OEFFNER, R. D., MILLÁN, C., SAMMITO, M., NASCIMENTO, A. J., USÓN, I. & READ, R. J. 2017. Gyre and Gimble in Phaser: a maximum likelihood replacement for PC-refinement. *Acta Crystallogr D Biol Crystallogr*.

METHA, P. K., HALE, T. I. & CHRISTEN, P. 1993. Aminotransferases: demonstration of homology and division into evolutionary subgroups.

. *Eur. J. Biochem*.

MILLÁN, C., SAMMITO, M., GARCIA-FERRER, I., GOULAS, T., SHELDRICK, G. & USÓN, I. 2015a. Combining phase information in reciprocal space for molecular replacement with partial models. *Acta Crystallographica Section D: Biological Crystallography*

, 71**,** 1931-1945.

MILLAN, C., SAMMITO, M., GARCIA-FERRER, I., GOULAS, T., SHELDRICK, G. M. & USON, I. 2015. Combining phase information in reciprocal space for molecular replacement with partial models. *Acta Crystallogr D Biol Crystallogr,* 71**,** 1931-45.

MILLÁN, C., SAMMITO, M. & USÓN, I. 2015b. Macromolecular ab initio phasing enforcing secondary and tertiary structure. *IUCrJ,* 2**,** 1:95.

MIURA, T. & KAJIMOTO, T. 2001. Application of L-threonine aldolase-catalyzed reaction to the preparation of protected 3R,5R-dihydroxy-L-homoproline as a mimetic of idulonic acid. *Chirality,* 13**,** 577-80.

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**,** 240-55.

NIRMALAN, N., WANG, P., SIMS, P. F. & HYDE, J. E. 2002. Transcriptional analysis of genes encoding enzymes of the folate

pathway in the human malaria parasite Plasmodium falciparum. *Mol Microbiol,* 46**,** 179-90.

O'BRIEN, P. J. & HERSCHLAG, D. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol,* 6**,** R91-R105.

OGAWA, H. & FUJIOKA, M. 2001. Purification and Characterization of Cytosolic and Mitochondrial Serine Hydroxymethyltransferases From Rat Liver. *J Biochem,* 90 381-390. 8.

OLANIRAN, A. O., PILLAY, D. & PILLAY, B. 2004. Haloalkane and haloacid dehalogenases from aerobic bacterial isolates indigenous to contaminated sites in Africa demonstrate diverse substrate specificities. *Chemosphere,* 55**,** 27-33.

OPPERMANN, U., FILLING, C., HULT, M., SHAFQAT, N., WU, X., LINDH, M., SHAFQAT, J., NORDLING, E., KALLBERG, Y., PERSSON, B. & JORNVALL, H. 2003. Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact,* 143-144**,** 247-53.

PAI, V. R., RAJARAM, V., BISHT, S., BHAVANI, B. S., RAO, N. A., MURTHY, M. R. & SAVITHRI, H. S. 2009. Structural and functional studies of Bacillus stearothermophilus serine hydroxymethyltransferase: the role of Asn(341), Tyr(60) and Phe(351) in tetrahydrofolate binding. *Biochem J,* 418**,** 635-42.

PAULING, L. 1946. Molecular Architecture and Biological Reactions. *Chem. Eng. News,* 24**,** 1375-1377.

PENNING, T. M. & JEZ, J. M. 2001. Enzyme redesign. *Chem Rev,* 101**,** 3027-46.

PERACCHI, A. 2001. Enzyme catalysis: removing chemically 'essential' residues by site-directed mutagenesis. *Trends Biochem Sci,* 26**,** 497-503.

PERCUDANI, R. & PERACCHI, A. 2003. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep,* 4**,** 850-4.

PLAMANN, M., STAUFFER, L., URBANOWSKY, M. & STAUFFER, G. 1983. Complete nucleotide sequence of E.coli glyA gene. *Nucleic Acids Res,* 11**,** 2065-2075.

RAJARAM, V., BHAVANI, B. S., KAUL, P., PRAKASH, V., APPAJI RAO, N., SAVITHRI, H. S. & MURTHY, M. R. 2007. Structure determination and biochemical studies on Bacillus stearothermophilus E53Q serine hydroxymethyltransferase and its complexes provide insights on function and enzyme memory. *FEBS J,* 274**,** 4148-60.

READ, R. 1990. Structure-factor probabilities for related structures. *Acta Crystallographica Section A: Foundations of Crystallography,* 46**,** 900-912.

READ R. J. , J, M. A. & C., S. L. 2007. Likelihood-based molecular replacement in phaser. *Evolving Methods for Macromolecular Crystallography. NATO Science Series,* 245.

READ, R. J. & MCCOY, A. J. 2016. A log-likelihood-gain intensity target for crystallographic phasing that accounts for experimental error. *Acta Crystallographica Section D,* 72**,** 375-387.

RENWICK, S. B., SNELL, K. & BAUMANN, U. 1998. The crystal structure of human cytosolic serine hydroxymethyltransferase: a target for cancer chemotherapy. *Structure,* 6**,** 1105-16.

ROZZELL, J. D. 1999. Commercial scale biocatalysis: myths and realities. *Bioorg Med Chem,* 7**,** 2253-61.

SCARSDALE, J. N., KAZANINA, G., RADAEV, S., SCHIRCH, V. & WRIGHT, H. T. 1999. Crystal structure of rabbit cytosolic serine hydroxymethyltransferase at 2.8 A resolution: mechanistic implications. *Biochemistry,* 38**,** 8347-58.

SCARSDALE, J. N., RADAEV, S., KAZANINA, G., SCHIRCH, V. & WRIGHT, H. T. 2000. Crystal structure at 2.4 A resolution of E. coli serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. *J Mol Biol,* 296**,** 155-68.

SCHALLMEY, M., FLOOR, R. J., HAUER, B., BREUER, M., JEKEL, P. A., WIJMA, H. J., DIJKSTRA, B. W. & JANSSEN, D. B. 2013. Biocatalytic and structural properties of a highly engineered halohydrin dehalogenase. *Chembiochem,* 14**,** 870-81.

SCHALLMEY, M., KOOPMEINERS, J., WELLS, E., WARDENGA, R. & SCHALLMEY, A. 2014. Expanding the Halohydrin Dehalogenase Enzyme Family: Identification of Novel Enzymes by Database Mining. *Appl Environ Microbiol,* 80**,** 7303-15.

SCHIRCH, D., DELLE FRATTE, S., IURESCIA, S., ANGELACCIO, S., CONTESTABILE, R., BOSSA, F. & SCHIRCH, V. 1993. Serine hydroxymethyltransferase: role of the active site lysine in the mechanism of the enzyme. *Adv Exp Med Biol,* 338**,** 715-8.

SCHIRCH, L. 1982. Serine hydroxymethyltransferase. *Adv Enzymol Relat Areas Mol Biol,* 53**,** 83-112.

SCHIRCH, V., HOPKINS, S., VILLAR, E. & ANGELACCIO, S. 1985. Serine hydroxymethyltransferase from Escherichia coli: purification and properties. *J Bacteriol,* 163**,** 1-7.

SCHIRCH, V., SHOSTAK, K., ZAMORA, M. & M., G.-B. 1991. The origin of reaction specificity in Serine Hydroxymethyltransferase. *J.Biol Chem,* 266**,** 759-764.

SCHIRCH, V. & SZEBENYI, D. M. 2005. Serine hydroxymethyltransferase revisited. *Curr Opin Chem Biol,* 9**,** 482-7.

SCHMID, E. F. & SMITH, D. A. 2002. Should scientific innovation be managed? *Drug Discov Today,* 7**,** 941-5.

SCHULZE, A. & HARRIS, A. L. 2012. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature,* 491**,** 364-73.

SIEGEL, R., MA, J., ZOU, Z. & JEMAL, A. 2014. Cancer statistics, 2014. *CA Cancer J Clin,* 64**,** 9-29.

SILVERMAN, R. B. 2002. The Organic Chemistry of Enzyme-catalyzed Reactions. *Academic Press.*

SINGH, R., SPYRAKIS, F., COZZINI, P., PAIARDINI, A., PASCARELLA, S. & MOZZARELLI, A. 2013. Chemogenomics of pyridoxal 5'-phosphate dependent enzymes. *J Enzyme Inhib Med Chem,* 28**,** 183-94.

SNELL, K. 1984. Enzymes of serine metabolism in normal, developing and neoplastic rat tissues. *Adv Enzyme Regul,* 22**,** 325-400.

SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res,* 33**,** W244-8.

SRIVASTAVA, H. K., BOHARI, M. H. & SASTRY, G. N. 2012. Modeling anti-HIV compounds: the role of analogue-based approaches. *Curr Comput Aided Drug Des,* 8**,** 224-48.

STITT, M. & GIBON, Y. 2014. Why measure enzyme activities in the era of systems biology? *Trends Plant Sci,* 19**,** 256-65.

STOVER, P. & SCHIRCH, V. 1991. 5-Formyltetrahydrofolate polyglutamates are slow tight binding inhibitors of serine hydroxymethyltransferase. *J Biol Chem,* 266**,** 1543-50.

STOVER, P. J. 2004. Physiology of folate and vitamin B12 in health and disease. *Nutr Rev,* 62**,** S3-12; discussion S13.

STOVER, P. J. & FIELD, M. S. 2011. Trafficking of intracellular folates. *Adv Nutr,* 2**,** 325-31.

SUMNER, J. 1926. The isolation and crystallization of the enzyme urease. *J. Biol. Chem,* 69.

SZEBENYI, D. M., MUSAYEV, F. N., DI SALVO, M. L., SAFO, M. K. & SCHIRCH, V. 2004. Serine hydroxymethyltransferase: role of glu75 and evidence that serine is cleaved by a retroaldol mechanism. *Biochemistry,* 43**,** 6865-76.

TANG, L., VAN MERODE, A. E., LUTJE SPELBERG, J. H., FRAAIJE, M. W. & JANSSEN, D. B. 2003. Steady-state kinetics and tryptophan fluorescence properties of halohydrin dehalogenase from Agrobacterium radiobacter. Roles of W139 and W249 in the active site and halide-induced conformational change. *Biochemistry,* 42**,** 14057-65.

TONEY, M. D. 2005. Reaction specificity in pyridoxal phosphate enzymes. *Arch Biochem Biophys,* 433**,** 279-87.

TOSCANO, M. D., WOYCECHOWSKY, K. J. & HILVERT, D. 2007. Minimalist active-site redesign: teaching old enzymes new tricks. *Angew Chem Int Ed Engl,* 46**,** 3212-36.

TRIVEDI, V., GUPTA, A., JALA, V. R., SARAVANAN, P., RAO, G. S., RAO, N. A., SAVITHRI, H. S. & SUBRAMANYA, H. S. 2002. Crystal structure of binary and ternary complexes of serine hydroxymethyltransferase from Bacillus stearothermophilus: insights into the catalytic mechanism. *J Biol Chem,* 277**,** 17161-9.

TUFVESSON P., LIMA-RAMOS J., NORDBLAD M. & M., W. J. 2011. Guidelines and Cost Analysis for Catalyst Production in Biocatalytic Processes. *Org. Process Res. Dev.,* 15 266-274.

VAN DEN WIJNGAARD, A. J., REUVEKAMP, P. T. & JANSSEN, D. B. 1991. Purification and characterization of haloalcohol dehalogenase from Arthrobacter sp. strain AD2. *J Bacteriol,* 173**,** 124-9.

VAN HYLCKAMA VLIEG, J. E., TANG, L., LUTJE SPELBERG, J. H., SMILDA, T., POELARENDS, G. J., BOSMA, T., VAN MERODE, A. E., FRAAIJE, M. W. & JANSSEN, D. B. 2001. Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *J Bacteriol,* 183**,** 5058-66.

VIDAL, L., CALVERAS, J., CLAPES, P., FERRER, P. & CAMINAL, G. 2005. Recombinant production of serine hydroxymethyl transferase from Streptococcus thermophilus and its preliminary evaluation as a biocatalyst. *Appl Microbiol Biotechnol,* 68**,** 489-97.

VLIEG, J., TANG, L., SPELBERG, J., SMILDA T., POELARENDS, G., BOSMA, T., VAN MERODE, J. F., M. & JANSSEN, D. 2001.

Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *JBacteriol,* 183**,** 5058-5066.

WANG, X., HAN, S., YANG, Z. & TANG, L. 2015. Improvement of the thermostability and activity of halohydrin dehalogenase from Agrobacterium radiobacter AD1 by engineering C-terminal amino acids. *J Biotechnol,* 212**,** 92-8.

WATANABE, F., YU, F., OHTAKI, A., YAMANAKA, Y., NOGUCHI, K., YOHDA, M. & ODAKA, M. 2015. Crystal structures of halohydrin hydrogen-halide-lyases from Corynebacterium sp. N-1074. *Proteins,* 83**,** 2230-9.

WILKINS, D. W., SCHMIDT, R. H. & KENNEDY, L. B. 1986. Threonine aldolase activity in yogurt bacteria as determined by headspace gas chromatography. . *J. Agric. Food Chem.,* 34**,** 150-152.

WINDLE, C. L., MULLER, M., NELSON, A. & BERRY, A. 2014. Engineering aldolases as biocatalysts. *Curr Opin Chem Biol,* 19**,** 25-33.

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., 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**,** 235-42.

WITTIG, U., REY, M., KANIA, R., BITTKOWSKI, M., SHI, L., GOLEBIEWSKI, M., WEIDEMANN, A., MULLER, W. & ROJAS, I. 2014. Challenges for an enzymatic reaction kinetics database. *Febs J,* 281**,** 572-82.

XANTHOPOULOS, V., PETRIDIS, D. & TZANETAKIS, N. 2001. Characterization and Classification of Streptococcus thermophiles and Lactobacillus delbrueckii subsp. Bulgaricus Strains Isolated from Traditional Greek Yogurts. *Journal of Food Science,* 66**,** 747-752.

YADAV, V. S., MISHRA, K. P., SINGH, D. P., MEHROTRA, S. & SINGH, V. K. 2005. Immunomodulatory effects of curcumin. *Immunopharmacol Immunotoxicol,* 27**,** 485-97.

YOSHIKUNI, Y., FERRIN, T. E. & KEASLING, J. D. 2006. Designed divergent evolution of enzyme function. *Nature,* 440**,** 1078-1082.

YU, F., NAKAMURA, T., MIZUNASHI, W. & WATANABE, I. 1994. Cloning of two halohydrin hydrogen-halide-lyase genes of Corynebacterium sp. strain N-1074 and structural comparison of the genes and gene products. *Biosci Biotechnol Biochem,* 58**,** 1451-7.

ZHANG, W. C., SHYH-CHANG, N., YANG, H., RAI, A., UMASHANKAR, S., MA, S., SOH, B. S., SUN, L. L., TAI, B. C., NGA, M. E., BHAKOO, K. K., JAYAPAL, S. R., NICHANE, M., YU, Q., AHMED, D. A., TAN, C., SING, W. P., TAM, J., THIRUGANANAM, A., NOGHABI, M. S., PANG, Y. H., ANG, H. S., MITCHELL, W., ROBSON, P., KALDIS, P., SOO, R. A., SWARUP, S., LIM, E. H. & LIM, B. 2012. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell,* 148**,** 259-72.