STRUCTURAL STUDIES OF ENZYMES INVOLVED IN PROPYLENE AND ACETONE METABOLISM IN *Xanthobacter autotrophicus*

by

Arathi Mandyam Krishnakumar

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

January 2007

APPROVAL

of a dissertation submitted by

Arathi Mandyam Krishnakumar

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Chair

Dr. John W. Peters

Approved for the Department of Chemistry and Biochemistry

Dr. David J. Singel

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S.Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeb Road, Ann Arbor, Michigan 48106, to whom I have granted “the exclusive right to reproduce and distribute my dissertation in and from microfilm along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part.”

Arathi Mandyam Krishnakumar

January 2007

ACKNOWLEDGEMENTS

I would like to thank my academic advisor Dr. John Peters for giving me the opportunity to do this work, for his guidance, patience, and financial support. Dr. Peters has been a very strong mentor to me these past few years and I have learned immensely from him – not only technically regarding this work, but how to approach work and research in general. His mentoring will influence me far beyond this work.

I would like to thank Drs. Martin Lawrence, Valérie Copié, Trevor Douglas, and Martin Teintze for their time, excellent technical guidance, encouragement and for serving on my thesis committee. Their role has been invaluable in the completion of this work. I would like to thank out collaborator Dr. Scott Ensign for his support and inputs into the project. I would also like to thank the staff at the Chemistry and biochemistry department – Jennifer, Chris, Sherry, Pete and Scott Busse for their timely help. My experience at MSU was made richer by my interactions with these people. Their encouragement, understanding, guidance, and friendship are deeply appreciated. I would like to thank the entire staff of Stanford Synchrotron Radiation Lab for their help and support. Last but furthest from the least, I would like to thank my parents, Anandalwar M.K and Rama Anand, my husband Ganesh Shankarling and my brother Sudarshan Krishnakumar, invaluable friends and family. Their contributions cannot be put in words, and I cannot thank them enough. I dedicate this dissertation to them.

TABLE OF CONTENTS

TABLE OF CONTENTS-CONTINUED

TABLE OF CONTENTS-CONTINUED

LIST OF TABLES

LIST OF TABLES-CONTINUED

LIST OF FIGURES

LIST OF FIGURES-CONTINUED

Figure Page

LIST OF FIGURES-CONTINUED

Figure Page

LIST OF FIGURES-CONTINUED

Figure Page

LIST OF FIGURES-CONTINUED

Figure Page

LIST OF FIGURES-CONTINUED

GLOSSARY

R-HPC	2-(R) Hydroxypropylthioethanesulfonate
S-HPC	2-(S) Hydroxypropylthioethanesulfonate
R-HPCDH	2-(*R*) Hydroxypropylthioethanesulfonate dehydrogenase
S-HPCDH	2-(S) Hydroxypropylthioethanesulfonate dehydrogenase
SDR	Short-Chain Dehydrogenase Reductase
2-KPC	2 Ketopropylthioethanesulfonate;
CoM	Coenzyme M (2-mercaptoethane sulfonate);
2-KPCC	2-Ketopropyl-CoM Oxidoreductase/Carboxylase;
MAD	Multiple-Wavelength Anomalous Diffraction;
NAD^+/NADH	Nicotinamide Adenine Dinucleotide;
$NADP^+$	Nicotinamide Adenine Dinucleotide phosphate;
MDH	Mannitol Dehydrogenase;
HADH	HaloAlcohol Dehydrogenase
MIR	Multiple Isomorphous Replacement
MR	Molecular Replacement
ATP	Adenosine Triphosphate
AMP	Adenosine Monophosphate
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
EaCoMT	Epoxyalkane CoM Transferase
MDR	Medium Chain Dehydrogenases Reductases
AMO	Alkene Monoxygenase
MMO	Methane Monoxygenase
Da	Daltons
kDa	Kilo Daltons
MPa	Mega Pascals
2'-IATP	2' Iodo-Adenosine Tri Phosphate
eV	Electron Volts
SSRL	Stanford Synchrotron Radiation Laboratory
FOM	Figure of Merit
NCS	Non-Crystallographic Symmetry
EMTS	Ethyl Mercury Thio Salicylate
DM	Density modification
AC	Acetone Carboxylase
PEG	Poly Ethylene Glycol
MgATP	Magnesium Adenosine Tri Phosphate
mM	Milli Molar
mm	Milli Meter
Å	Armstrong
PEP	Phopshoenol Pyruvate

ABSTRACT

X-ray crystallography has been an indispensable tool in understanding the mechanism of the enzymes of the epoxide carboxylation pathway in *Xanthobacter autotrophicus*. The main focus of this dissertation involves providing the structural basis for the stereoslectivity of the two stereospecific dehydrogenases of the pathway namely *R*- and *S*-HPCDH. The crystal structure of *R*- HPCDH co-crystallized with the substrate has been determined. The key elements of interactions between the enzyme and substrate are electrostatic interactions between the sulfonate oxygen atoms and two arginine residues (Arg152 and Arg196) of *R*- HPCDH. The comparison of the structure of *R*- HPCDH with a homology model of the *S*-HPCDH provides a structural basis for a mechanism of substrate specificity in which the binding of the substrate sulfonate moiety at distinct sites on each stereoselective enzyme directs the orientation of the appropriate substrate enantiomer for hydride abstraction. Moreover, crystal structures of the two methionine mutants of *R*-HPCDH have revealed that they have a role in shielding electrostatic interactions between the enzyme and the substrate from the surroundings. The structure of the presumed product bound state reveals that binding interactions between the substrate and the enzyme have striking similarities to the ones observed in the previously determined structure of 2-KPCC highlighting the utility of coenzyme M as a carrier molecule in the pathway. Extensive comparative structural analyses of the enzymes of the pathway reveal a common structural signature for coenzyme M binding. Coenzyme M, when conjugated with the substrates that lack innate chemical groups, such as short chain alkenes and epoxides, provide these substrates a handle to specifically bind and interact with the enzymes, thereby orienting them in a proper fashion for catalysis. Finally, exhaustive amount of ground work is laid towards the determination of the three dimensional structure of acetone carboxylase, the enzyme which converts acetone to acetoacetate in Xanthobacter. Initial electron density maps calculated by the phase information obtained from a number of moderately good heavy atom derivatives show a clear protein-solvent envelope, providing the first glimpse into the three dimensional structure of acetone carboxylase.

CHAPTER1

INTRODUCTION

Industrialization at a rapid pace has paved the way to increasing concentrations of a variety of carbon-based gases, including CO_2, CO, aliphatic and aromatic hydrocarbons, and halocarbons. Many alkenes are also produced as natural intermediates or end products of bacterial, plant and mammalian metabolism. Vinyl chloride (VC) is a common ground water contaminant [1] which is of concern due to its carcinogenicity [2]. Although it can be produced naturally at very low concentrations in some soils [1, 3], the industrial synthesis of polyvinyl chloride plastics [2] and the bacterial metabolism of chlorinated solvents [4] are the most problematic sources of VC contamination. Other alkenes like propylene and its oxide are widely used in the industrial synthesis of propylene glycol and polyurethane foam and are major sources of contamination. The toxicity of aliphatic alkenes arises from their ability to readily serve as substrates for broad specificity oxygenases such as cytochromeP450 enzymes resulting in the formation of epoxides. It is these epoxides which pose a bigger problem than the alkenes because of their high biological reactivity.

The Biological Reactivity of Epoxides

Epoxides and ketones are widely used in industry and have been shown to have potential mutagenic, toxic and carcinogenic effects [5]. The biological

reactivity of epoxides is depicted in Figure1. Epoxides are formed by epoxidation of alkenes by the enzymes with monoxygenase activity. For example cytochromeP450 catalyzes the epoxidation of alkene by using it as a substrate in place of a natural oxygenase substrate [6]. Once formed, the epoxides are highly reactive and can have various fates with respect to their biochemistry [7]. Epoxides, being electrophiles, can undergo reaction with cellular nucleophiles such as proteins and DNA and form covalent adducts, thereby leading to mutations and cancer.

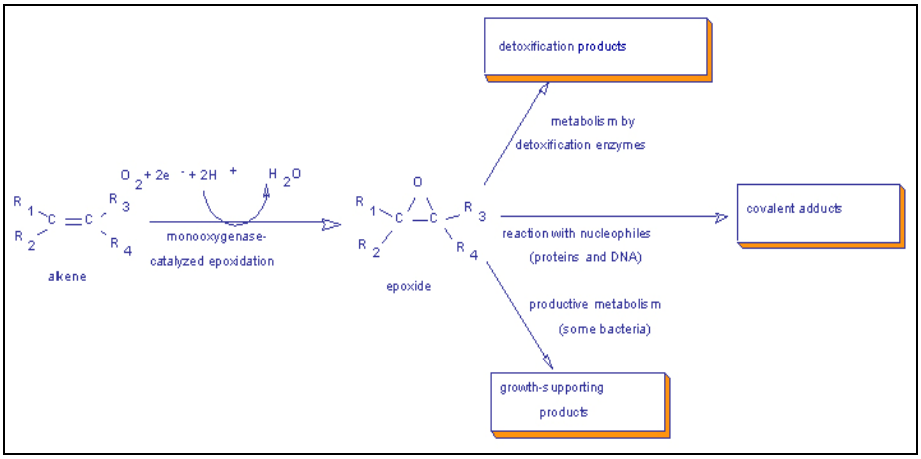

Figure 1. Biological reactivity of epoxides.

To counteract these kind of reactions some organisms utilize detoxifying enzymes such as glutathione-S-tranferases [8] and epoxide hydrolases [9] that use glutathione or water, respectively, as nucleophiles attack the oxirane ring and opens them to form alcohols (equations 1 and 2, respectively in Figure 2). These products may be either metabolized further or discarded as waste products of metabolism.

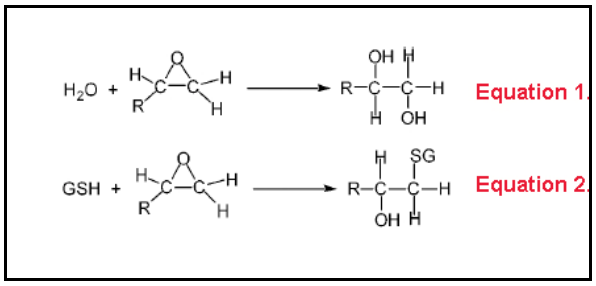

Figure 2. Reactions of epoxide metabolism by epoxide hydrolase and Glutathione S-transferase depicted in equations 1 and 2, respectively.

Alkene Oxidizing Microorganisms

Diverse microorganisms including various strains of R*hodococcus, Mycobacterium, Nocardia, Xanthobacter, Alcaligenes, Pseudomonas, Ochrobactrum, Nocardiodes and Corynebacterium*, are capable of growth under aerobic conditions utilizing short-chain aliphatic alkenes (C2-C6) as the sole source of carbon [7, 10]. These organisms can utilize alkenes like ethene, propylene, 1-butene, butadiene, isoprene, vinyl chloride, etc. While most of these microorganisms are restricted to growth on a single alkene, *Xanthobacter autotrophicus* strain Py2 is capable of growing on a wide range of substrates. Originally isolated with propylene, this bacterium is capable of utilizing ethylene, 1-butylene, 2-butylene, 1-pentene, and 1-hexene as carbon sources [11, 12]. An additional feature of *Xanthobacter* is that it is capable of growing on alternative substrates like glucose, fructose, organic acids, C-1 to C-4 alcohols, acetone, 1, 2-propanediol and a mixture of H_2 and CO_2 [11]. Propylene-grown *Xanthobacter* cells (strain Py2) also degrade several chlorinated alkenes of environmental concern, including trichloroethylene, 1-chloroethylene (vinyl

chloride), cis- and trans-1,2-dichloroethylene, 1,3-dichloropropylene, and 2,3-dichloropropylene [13]. All microorganisms oxidizing alkenes isolated to date are aerobic and initiate alkene oxidation by a monoxygenase reaction. This is the first reaction where the alkene is converted to an epoxide and is carried out by O2- and reductant-dependent monoxygenases called alkene monoxygenases (AMO). Alkene monoxygenases are discussed in detail later in this chapter.

One of the most common mechanisms that bacteria employ to metabolize epoxides is carboxylation of these compounds by utilizing novel carboxylases [14]. Although carboxylases are diverse in terms of their structure, substrate specificity, and cofactor usage, they all have a common feature with regard to their ability to generate a stabilized carbanion for attack on electrophilic CO*2* or activated CO_2 species [15]. Different carboxylases use different chemistry to catalyze the carboxylation reaction. Ribulose 1, 5-diphosphate carboxylase (Rubisco), uses general acid-base chemistry along with metal ion catalysis to form the cis enediolate intermediate [16, 17]. Metal ion catalysis is also used in the well known enzyme of the glycolysis pathway phosphoenol pyruvate carboxylase where the reaction is catalyzed by the formation of the enolate tautamer of pyruvate carbanion when bicarbonate attacks and removes the phosphate group of PEP [17]. Another example of acid-base catalysis is seen in the vitamin K-dependent carboxylases where oxygenation of vitamin K facilitates the formation of a strong base which in turn abstracts a proton from the γ-carbon of glutamate [18]. In all these cases it is interesting to note that the

substrate itself is a potential nucleophile and the proton abstraction is carried out by the enzyme to generate the carbanion. Contrary to this general scheme, there are carboxylases with unique molecular properties and cofactor requirements which convert aliphatic epoxides to β-keto acids [13, 19, 20]. Epoxides themselves are strong nucleophiles and they are carboxylated via unique transformations to the acid product by novel carboxylation reactions.

To date the best characterized pathway for alkene metabolism is that of propylene in *Xanthobacter autotrophicus*. As mentioned earlier, the metabolism of propylene starts by its conversion to the *R* and *S* enantiomers of epoxypropane. The enzymes and the reaction mechanisms involved in this pathway are introduced and discussed in detail in the following sections of this chapter.

Alkene Monoxygenases (AMOs)

The most characterized alkene monoxygenases are from two propylene-oxidizing bacteria: *Rhodococcus rhodochrous* strain B276 [21] and *Xanthobacter autotrophicus* strain Py2 [22]. These enzymes are multi component enzymes that catalyze epoxidation of alkenes in an NADH-dependent fashion as shown in Figure 3.

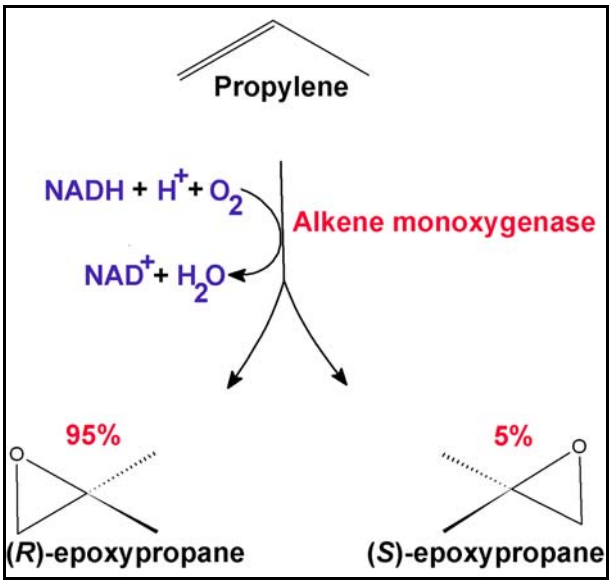

Figure 3. NADH-dependent epoxidation of propylene by Alkene monoxygenase.

The multi component AMO enzyme system of *R. rhodochrous* and *X. autotrophicus* are shown in Figures 4 and 5, respectively. AMO from *R. rhodochrous* is a three-component enzyme system consisting of an iron containing epoxygenase which has αβ quaternary structure composed of two subunits of 53 and 35-kDa and a 40kDa monomeric reductase that contains FAD and a 2Fe-2S cluster, and a 14 kDa effector protein (Figure 4).

Based on the evidence from biochemical and spectroscopic studies, this enzyme has been classified into the class of diiron oxygensases [23-25]. The best characterized enzyme of this family is the soluble methane monoxygenase (MMO). Crystal structure of the hydroxylase component of MMO reveals that the enzyme contains a binuclear iron center with two terminal histidine and glutamate residues, two bridging glutamate residues and water molecule(s) forming a bis-μ-hydroxo-bridged "diamond core" structure [26, 27].

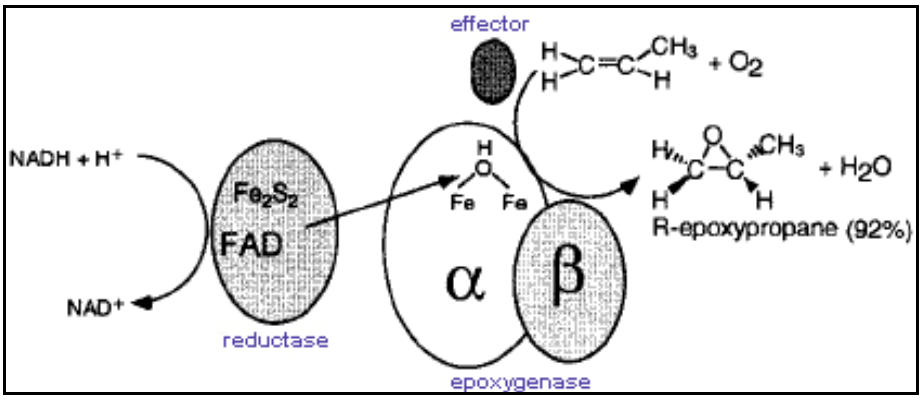

Figure 4. Multi component AMO enzyme system of *R. rhodochrous*.

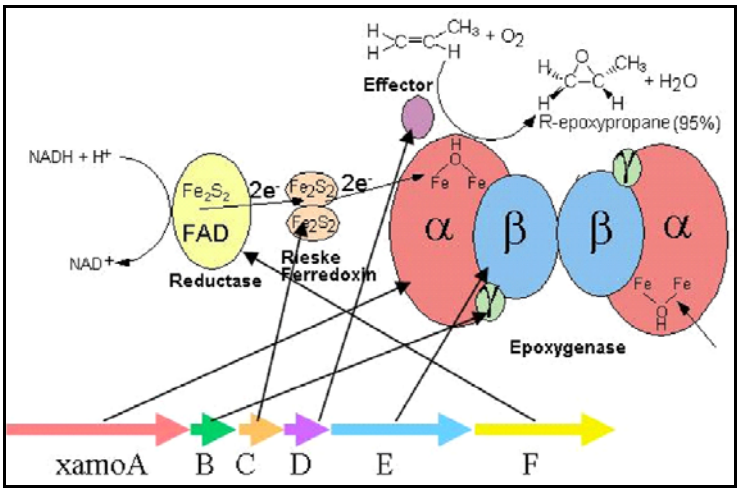

Figure 5. Multi component AMO enzyme system of *X. autotrophicus.*

The epoxidation reaction of AMO and the hydroxylation reaction of MMO are carried out at the diiron centers. The binding of oxygen and the activation of organic substrate occurs at this site. The other monoxygenase which has been purified to homogeneity and well characterized both biochemically and spectroscopically is the AMO from *X. autotrophicus*. Although AMO from *X. autotrophicus* is similar to AMO from *R. rhodochrous* with regard to the presence of diiron catalytic center, it is very different with regards to the quaternary structure. AMO from *X. autotrophicus* is composed of four rather

than three components (Figure 5). The first component is an epoxygenase containing four moles of iron and consists of three subunits of 38, 58 and 10 kDa, respectively arranged in a α2β2γ2 quaternary structure. The second component is a 35 kDa monomeric reductase that contains FAD and a 2Fe-2S cluster. The third and the fourth components are a homodimeric ferredoxin (13kDa) containing two Rieske-type 2Fe-2S clusters and an 11kDa effector protein, respectively [28].

As in the case of *R. rhodochrous* AMO, biochemical and genetic characterization have revealed that iron is arranged in a binuclear MMO-like cluster bound to the alpha subunit of the epoxygenase. As opposed to broad-range oxygenases which are not stereo selective, AMOs exhibit a high degree of stereo selectivity for alkene epoxidation. Purified AMOs from *X. autotrophicus* strain Py2 and *R. rhodochrous* catalyze the oxidation of propylene with 95 and 92% yields, respectively, of *R*-epoxypropane [25, 29].

The Pathway of Epoxide Carboxylation

The best characterized pathway of alkene metabolism is the conversion of propylene to acetoacetate in *X. autotrophicus* strain Py2 [25, 29-34]. As shown in Figure 6, the pathway consists of three steps, catalyzed by four enzymes. The pathway requires CO_2, NAD^+, NADPH, and an unusual cofactor, coenzyme M/2-mercaptoethane sulfonic acid [26]. The pathway has two interesting features: an unprecedented metabolic conversion of an epoxide to β-ketoacid and the

involvement of coenzyme M as a cofactor in the reaction [30]. Previously, coenzyme M was thought to be present and involved in methanogenesis in methanogenic archaea, where a methyl group is transferred to CoM to form an ethyl thioether that is subsequently reduced to methane [35-37]. The first step of the epoxide carboxylation pathway is the transfer of coenzyme M to form a 2-hydroxypropylthioether conjugate. This step is catalyzed by the enzyme Epoxypropane: coenzyme M transferase (EaCoMT) which can use either *R* or *S*-epoxypropane as a substrate forming the corresponding *R*- or *S*-hydroxypropyl thioether (hydroxypropylcoenzymeM) [30]. In the next step, two stereo specific dehydrogenases, *R* and *S*-hydroxypropylcoenzymeM dehydrogenases (*R*-HPCDH and *S*-HPCDH, respectively) catalyze the NAD^+ dependant dehydrogenation of *R* and *S*-hydroxypropylcoenzymeM (*R*- or *S*-HPC) to a common achiral product, 2-ketopropylcoenzymeM (2-KPC) [29].

In the final step, a novel carboxylase, 2-ketopropyl carboxylase/reductase (2-KPCC) catalyzes the NADPH-dependent reduction, cleavage and carboxylation of 2-KPC, forming acetoacetate and CoM. CoM is thus unchanged and recycled for the next round of reactions [38].

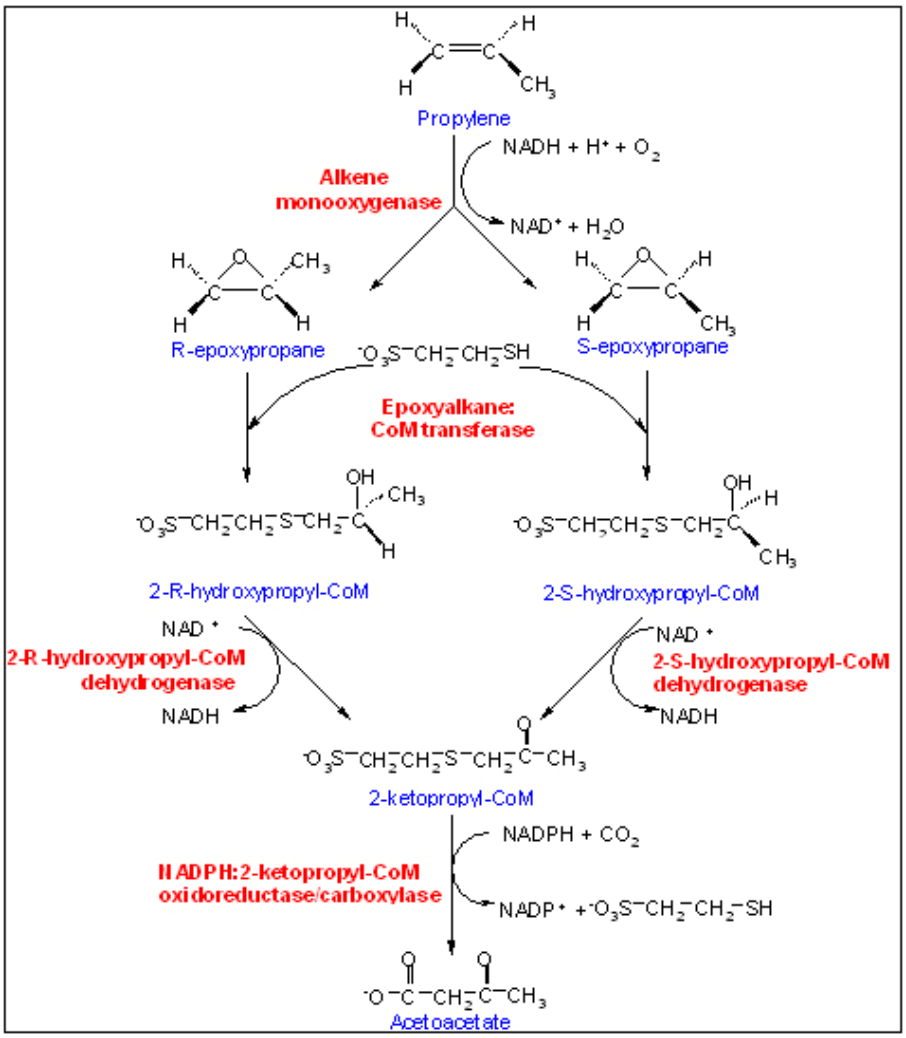

Figure 6. Epoxide carboxylation pathway in *X. autotrophicus* Epoxypropane.

The Enzymes of the Propylene Epoxidation Pathway

The genes encoding the four enzymes of the epoxidation have been cloned and sequenced in *X. autotrophicus* Py2 [39]. Analysis of amino acid sequences of the *X. autotrophicus* enzymes, together with biochemical, spectroscopic, and kinetic studies of the purified enzymes has helped in forming working models for

the mechanisms of these enzymes. Below is the summary of the models for each of the enzymes involved in the pathway.

Epoxypropane; CoM Transferase

EaCoMT catalyzes the reaction of CoM and epoxypropane involving the epoxide ring opening and conjugation to CoM resulting in the formation of a 2-hydroxypropylthioether (Figure 7). The transferase is not highly stereo selective, such that either the *R*-or *S*- enantiomer of epoxypropane formed upon alkene monoxygenase catalyzed epoxidation can be utilized as a substrate [30]. As purified, epoxypropane CoM transferase is an α6 hexameric protein (~250000 kDa) that contains one zinc atom per subunit. The enzymatic mechanism is believed to involve a zinc-based thiol activation, with direct analogy to that involved in the reaction of the cobalamin-independent methionine synthase (MetE) [32, 39, 40].

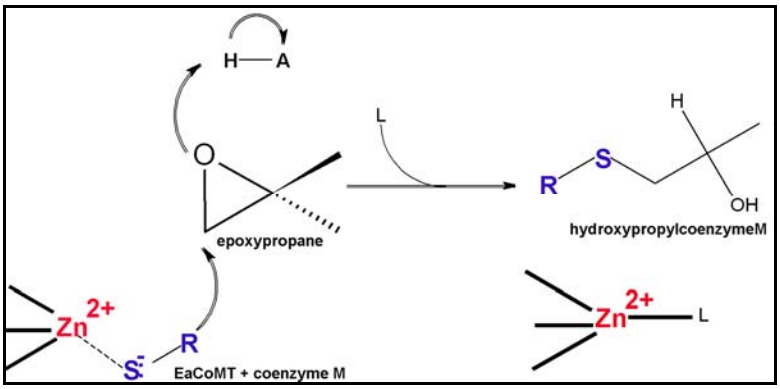

Figure 7. Reaction mechanism of transfer of coenzyme M to epoxypropane by EaCoMT involves a Zinc-activated thiol.

The similarities between the epoxypropane:coenzymeM transferase and the methionine synthases are distinguishable by amino acid sequence comparisons of the transferase to various cobalamin-independent methionine synthases from several sources [39]. These observations indicate that epoxypropane: coenzyme M transferase is a member of a family of enzymes that activate the thiol groups by coordination to zinc. Members of this family include farnesyl transferase [41, 42], the E.coli ADA protein [43, 44], S-methyl-methionine: homocysteine methyltransferase[45], betaine:homocysteine S- methyltransferase [46], and methylcob(III)almin;CoM methyltransferase (Mta) [47]. In all these enzymes the zinc is believed to coordinate the sulfur of the substrate, which in the case of EaCoMT is the sulfhydryl group of CoM thereby lowering the pKa of the thiol and activating it for alkylation. In all cases, three permanent ligands are donated from Cys and His residues, with the fourth coordination site occupied by an exchangeable ligand that is removed upon thiol binding. Recently there has been direct spectroscopic evidence for the role of the zinc ion of EaCoMT in CoM binding and activation for nucleophilic attack on the second substrate epoxypropane [48]. When the apo EaCoMT was reconstituted with Co^{2+} for spectroscopic studies it gave a normal spectrum corresponding to the tetrahedral state of the metal. Upon CoM addition to the enzyme, the spectrum was altered dramatically probably indicating that the thiol of CoM directly ligates the Co^{2+} center, either by displacing an exchangeable ligand or by increasing the coordination number about Co^{2+} [49]. There are no structural studies as yet on

EaCoMT. However, considering the high degree of homology of this enzyme to other enzymes of the family of thiol activated enzymes such as MetE, homology models can be constructed for this enzyme which will provide information about the substrate binding site and coenzyme M binding mode. These studies are discussed in detail in chapter 4 of this dissertation.

R- and *S*- Hydroxypropyl Coenzyme M Dehydrogenases

The *R*- and *S*- hydroxypropylcoenzyme M dehydrogenases (*R*- and *S*-HPCDH, respectively) catalyze the oxidation of 2-hydroxypropyl-CoM enantiomers to yield 2-ketopropyl-CoM (2-KPC). The two enzymes are highly specific for their respective substrates, exhibiting only 0.5 to 1% activity with the opposing enantiomer [29]. *R*- and *S*-hydroxypropyl coenzymeM dehydrogenases exist with subunit molecular weights of 25,400 and 26,000, respectively. Mechanistic details and amino acid sequence comparisons have revealed that these enzymes belong to the short-chain dehydrogenase/reductase (SDR) family of enzymes [50]. The *R*- and *S*-HPCDH enzymes share a relatively high degree of sequence identity (41%) and sequence similarity. *R*-HPCDH has been cloned and expressed in a fully active form, but unfortunately, *S*-HPCDH has not been successfully expressed due to its tendency to form inclusion bodies in heterologous expression systems [51] and have hampered further studies. Kinetic and mechanistic studies on *R*-HPCDH have shown that the enzyme follows a pattern consistent with the other members of the SDR family of

enzymes. The mechanism is ordered with NAD^+ binding first followed by the enantiomeric substrate *R*-HPC. It exhibits a K_{cat} of 26 s^{-1} and K_m for 2-*R*-hydroxypropylcoenzymeM (*R*-HPC) of 105 μM [51].

SDR Family of Enzymes

As mentioned earlier, *R*- and *S*-hydroxypropyl CoM dehydrogenases belong to the SDR family of enzymes. SDR enzymes catalyze the NAD(P)(H)-dependent dehydrogenation of a wide variety of substrates like alcohols, steroids, retenoids, prostaglandins, sugars and other small molecules. Dehydrogenases of this class are approximately 250 amino acids in length and exist as dimers and tetramers in solution. SDR enzymes constitute one of the largest super families of enzymes with ~1600 enzymes characterized to date [52, 53]. Enzymes within this family were initially distinguished from insect alcohol dehydrogenases and bacterial ribitol dehydrogenase based on their structural and mechanistic differences from other dehydrogenases. The amino acid sequence comparisons of the many SDR enzymes have suggested that they follow a distinctive mechanistic pattern and have paved the way to identifying them as a new family of enzymes.

Primary Structure of SDR Enzymes

Sequence alignment of enzymes belonging to the SDR family of enzymes reveals that despite low sequence identity (15-30%), there are two short consensus segments (Figure 8). The strictly conserved segment of YXXXK

sequence is the active-site motif. In addition to the Tyr and Lys, a serine residue which is 13 residues upstream to Tyr has been shown to be conserved in some SDR enzymes. Another sequence is the N-terminal GXXXGXXG sequence, which is similar to the GXGXXG(A) sequences characteristic of coenzyme binding fold in enzymes of other families [54]. In addition, an Asp residue is located between the βB and αC of most NAD^{+}(H)-dependent enzymes which is not conserved in $NADP^{+}$(H)dependent enzymes [55].

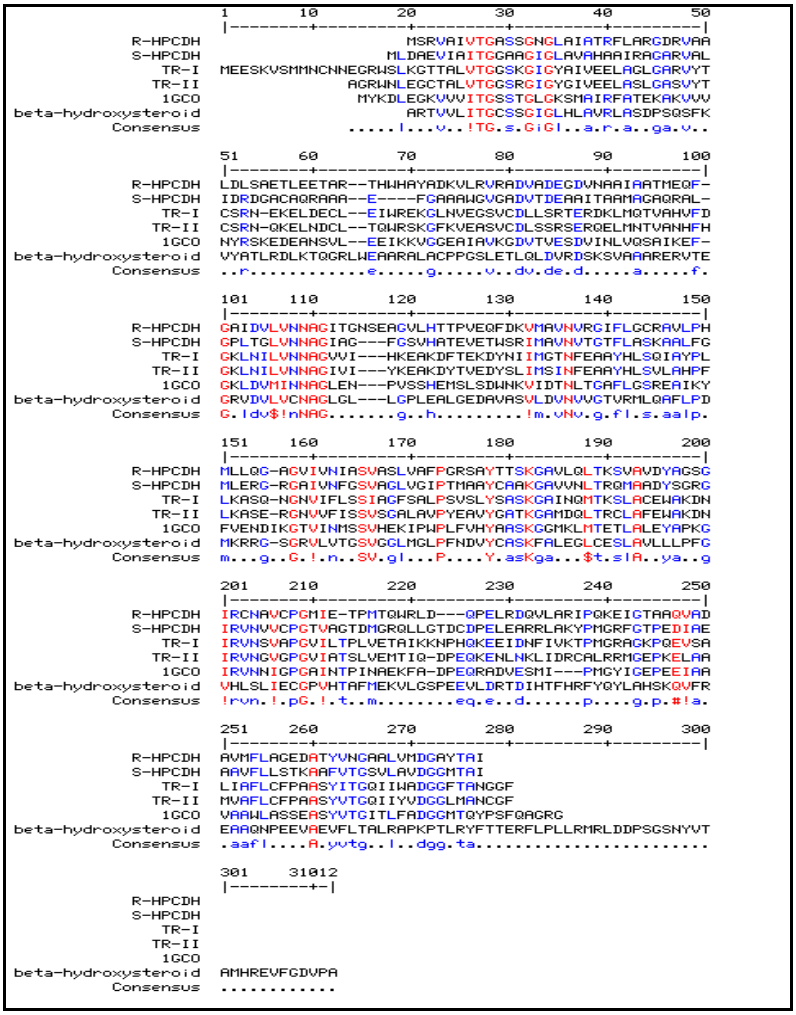

Figure 8. Sequence alignment of some of the SDR enzymes showing conserved sequences in the nucleotide binding region as well as the active site.

Tertiary Structure of the Enzymes of SDR Family

To date more than 15 structures of SDR enzymes have been determined. It is noteworthy that even though the primary sequence similarities are low among these enzymes, the tertiary structures are highly conserved. These structures have revealed that SDR enzymes are typically composed of two or four identical subunits that do not use any metal for catalysis. The overall structure of a subunit of a SDR enzyme is shown in Figure 9. It is an α/β type structure with an open twisted parallel 7-stranded β sheet in the middle surrounded by 6 α helices. Each subunit consists of two βαβαβ motifs; βA-αB-βB-αC-βC and βD-αE-βE-αF-βF exhibiting right-handed crossover [56] by a seventh β strand with a left-handed cross over connection between β strands F and G. The subunits are folded into three distinct domains: the N-terminal nucleotide binding domain, the central catalytic domain and the C-terminal substrate binding domain (Figure10). The N-terminal nucleotide binding domain consists of the classic Rossmann fold and the nucleotide binding sequence GGXXGXXG.

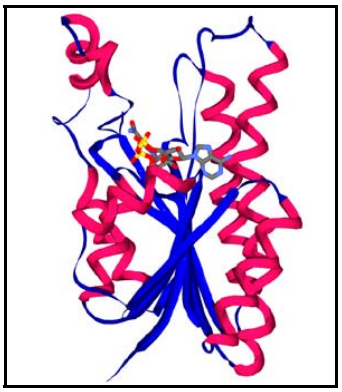

Figure 9. General tertiary structure of SDR enzymes. NAD+ is shown bound.

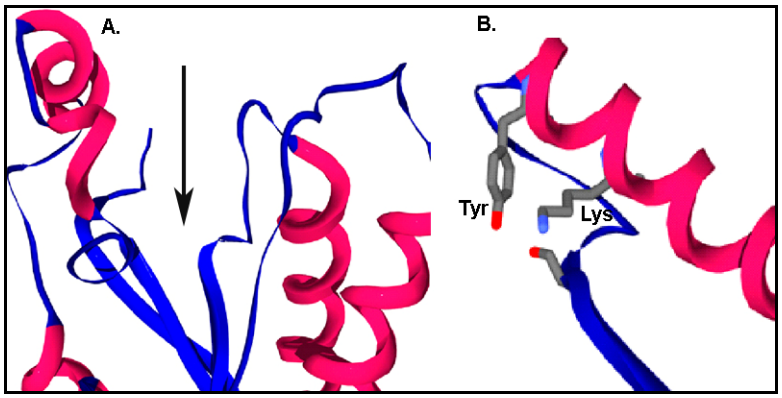

Figure 10. The substrate binding domain (A) of SDR enzyme is shown. Arrow indicates the substrate binding cleft. B indicates the catalytic triad in the central domain.

Several other Glycine residues are also conserved, as can be seen in the multiple sequence alignment (Figure 8), indicating their importance in the proper folding of the enzymes [56]. The central catalytic domain is home to the catalytic conserved sequence YXXXK. The substrate binding domain of various SDR enzymes has a conserved structure in different SDR enzymes although the primary sequences are different. The substrate binding domain is generally composed of the loops between the last two β strands (generally βF and βG) of a subunit on one side and the loop between the previous two β strands (βD and βE) on the other side [57].

General Mechanism of SDR Enzymes

The culmination of structural, kinetic and site directed mutagenesis studies on SDR enzymes has resulted in the formulation of a general mechanism of action [52, 53] (Figure11).

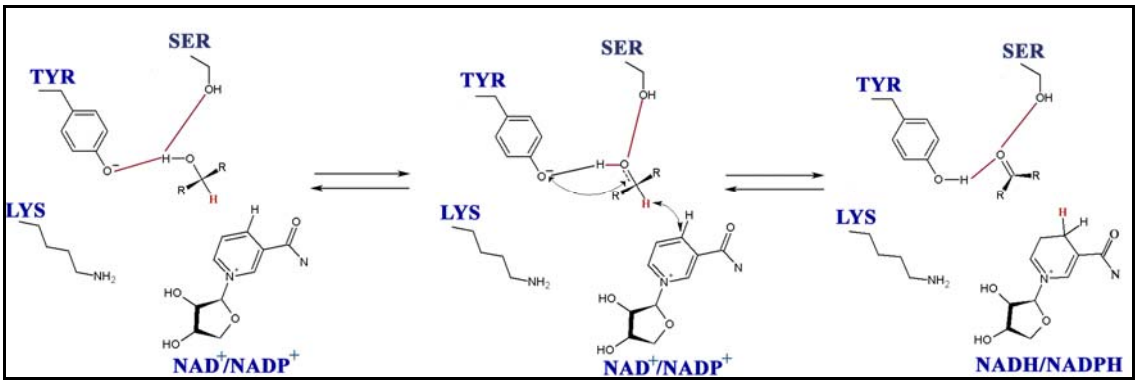

Figure 11. General catalytic mechanism of SDR enzymes.

In the first step of the reaction, the deprotonated phenolic group of tyrosine is hydrogen bonded to the hydroxyl group of the substrate. The serine residue of the catalytic triad also hydrogen bonds to the alcohol group thereby decreasing its pKa. In the second step, the deprotonated tyrosyl acts as a general base and abstracts a proton from the hydroxyl group of the substrate. Concomitantly, hydride transfer takes place to the B-face of the NAD^+ or $NADP^+$. The products of the reaction are a corresponding ketone of the alcohol and NADH/NADPH. The optimum pH for the dehydrogenation reaction by SDR enzymes is between 8 and 9 [58]. Since the reaction calls for a deprotonated tyrosine residue from the beginning of the reaction, the unusually low pKa of tyrosine may be achieved only by the presence of a positively charged residue nearby, which in this case is the catalytically conserved lysine residue. The lysine residue also has an important role in stabilizing the binding of the $NAD^+/NADP^+$ through a bifurcated hydrogen bond to both the 2'-hydroxyl and 3'-hydroxy group of the nicotinamide ribose.

Coenzyme Specificity in SDR Enzymes

Based on the coenzyme specificity, SDR enzymes can be divided into two groups that use either NAD(H) or NADP(H). The coenzyme specificity is generally related to the function of the enzymes, i.e., whether they act as reductases or dehydrogenases, because the intracellular ratio of NADPH to $NADP^+$ is high and conversely that of NADH to NAD^+ is significantly low. When a structure based alignment of SDR was carried out, the alignment revealed that there are two conserved stretches of amino acids responsible for cofactor specificity, the loop between βA and αB, and the loop between βB and αC [59]. The first loop has the GXXXGXG motif, where X can be any amino acid. In the enzymes using $NADP^+$, the position in the front of the second glycine is occupied by a positively charged amino acid (arginine or lysine), whereas enzymes using NAD^+ prefer small or even negatively charged amino acids(serine, threonine or aspartate). A second highly conserved arginine residue is located in the loop between βB and αC in enzymes utilizing $NADP^+$. This position is seen to be occupied by different amino acids in different enzymes utilizing NAD^+ [60].

Although to date greater than 1600 NAD^+ and $NADP^+$ -dependent enzymes have been reported and more than 60 of these are characterized biochemically to some degree including at least 15 structurally characterized, only a few have been discovered that have a homologous partner catalyzing the same reaction but with an opposite stereo selectivity. These enzymes are limited

to a set of plant tropinone reductases and the *R*- and *S*- hydroxypropylcoenzyme M dehydrogenases from *Xanthobacter autotrophicus* and *Rhodococcus rhodochrous*. Tropinone reductase I and II (TR-I and TR-II) from *Datura stramonium* catalyze the reversible stereo specific reduction of the 3-keto group of tropinone to form the enantiomers tropine (α-hydroxy) and pseudotropine (β-hydroxy), respectively [61]. Crystal structures of TR-I and TR-II have revealed the mechanism of stereo selectivity. According to this mechanism, electrostatic interaction between substrates and charged amino acids in the active site were responsible for orienting the substrate to its correct geometry [62, 63]. In TR-I, there a positively charged histidine (His-112) is located which repels the positively charged nitrogen on the tropinone (Figure 12). In the case of TR-II, the negatively charged glutamate carboxylate (Glu-156) attracts the substrate nitrogen and thus fixes its position [64, 65] (Figure 12). Based on these results it is conceivable that a similar charge attraction/repulsion mechanism, involving interactions with the negatively charged sulfonate moiety of the coenzymeM, controls proper binding of *R*- and *S*-2-hydroxypropyl-CoM to the respective dehydrogenases. In SDR enzymes it is known that the substrate binding domains are in the C-terminal region. Examination of the C-terminal sequences of the R and S dehydrogenases show similar sequence topologies interspersed with non identities (Figure 13). Several of these non identities involve oppositely charged amino acids. For example, a positive Arg residue at position 203 in the *R*-HPCDH has a negatively charged Glu counterpart in the *S*-HPCDH. Similarly,

an Asp residue at position 252 in *R*-HPCDH has a lysine counterpart in *S*-HPCDH. It has been speculated that charge differences at one or more of these positions provide a charge attraction/repulsion mechanism for orientation of the proper enantiomeric substrate. Amino acid sequence alignment of *R* and *S*-HPCDH has shown that these enzymes have a number of differences in amino acid sequence in their respective C-terminal domains, most striking being the difference in placement of arginine residues that may play a role in binding of the sulfonate of CoM through salt bridges (Figure14) [66]. Mutagenesis of these amino acids has suggested that Arginine 152 and 196 are crucial for substrate binding. For the mutant Arg152Ala and Arg196Als, kcat values for the natural substrate 2-KPC was decreased substantially and km values were drastically increased by 100-300 fold [67].

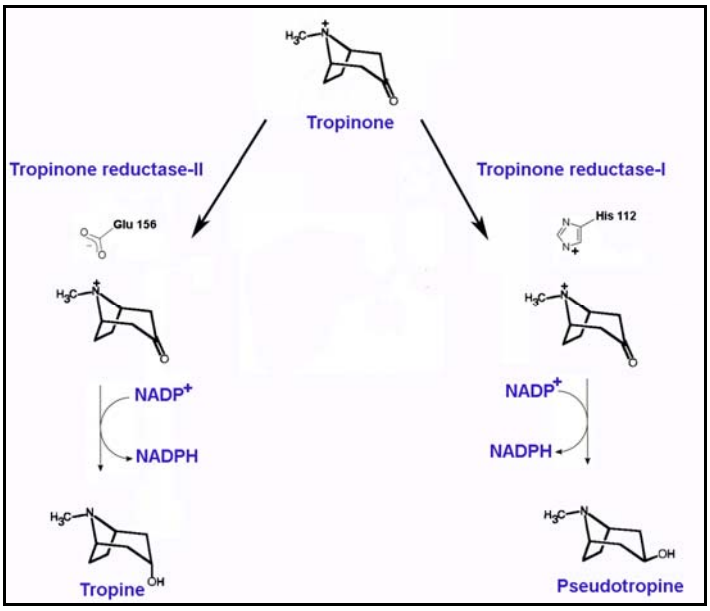

Figure 12. Charge/repulsion attraction mechanism in TR-I and TR-II for stereo selective catalysis of dehydrogenation of tropine and pseudotropine.

Figure 13. Global sequence alignment of C-terminal regions of *R*- and *S*-HPCDH. Arrows indicate non-identities which are discussed in text.

These results collectively suggest that interactions between the sulfonate of the CoM and specific arginine residues are the key to enantioselectivity and catalytic efficiency of *R*-HPCDH. A model is proposed where differential placement of arginine residues at the substrate binding site in the two enzymes brings about enantioselectivity and substrate specificity by controlling the specific interaction between the substrate sulfonate group and enzyme, thereby orienting the substrate correctly for hydride abstraction (Figure 14).

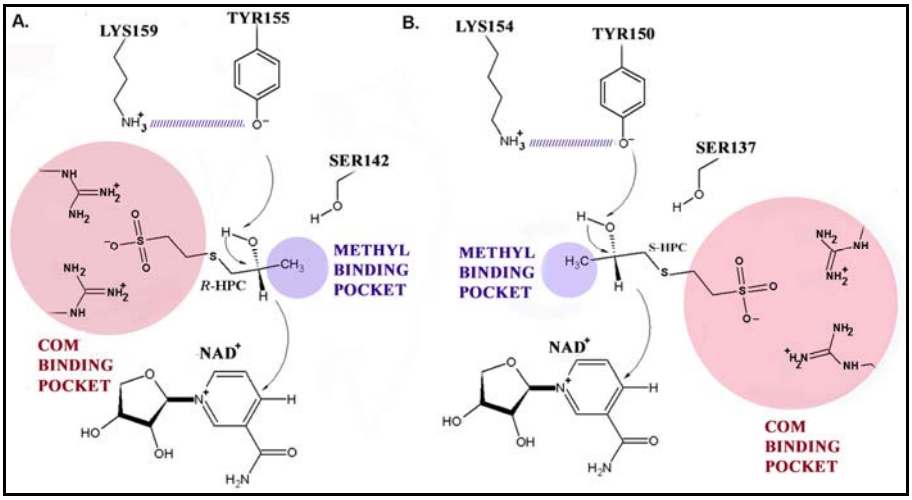

Figure 14. Models for substrate binding to (A) R-HPCDH and (B) S-HPCDH showing the involvement of differentially placed arginine residues in the CoM binding pocket.

Enzymes such as those that exhibit high degree of stereo selectivity have been studied extensively both in the context of elucidating the features that control the

substrate specificity and for utilization as biocatalysts in stereo selective synthesis. Since enantiopure chemicals are of high demand in the pharmaceutical industry it is very important to study the mechanism of stereo selectivity in these enzymes.

NADPH: 2-Ketopropyl-CoM Carboxylase/ Oxidoreductase (2-KPCC)

The final reductive cleavage of 2-ketopropyl CoM to acetoacetate and coenzyme M, which is the actual step where CO_2 is fixed, is catalyzed by the NADPH dependent enzyme 2-KPCC. 2-KPCC is a homodimeric enzyme consisting of 57kDa subunits. Each subunit has one molecule of FAD bound to it. Amino acid sequence analysis of 2-KPCC has showed that the enzyme belongs to the family of NADPH: disulfide oxidoreductase (DSOR) enzymes. Well known members of this family of enzymes are glutathione reductases, dihydrolipoamide dehydrogenase, and mercuric reductase. Apart from the fact that they are all NADPH-dependent and FAD-containing enzymes, an important feature of these enzymes is that they contain a highly conserved redox active cysteine pair that participates in substrate reduction. Biochemical and structural characterization of these enzymes have allowed formulation of a general mechanism of DSOR enzymes as shown in Figure15. .

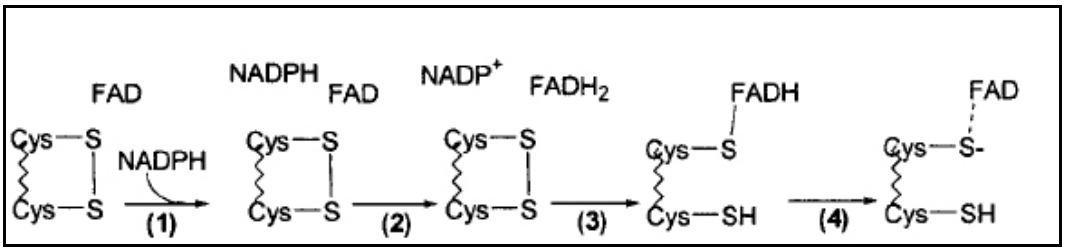

Figure 15. General mechanism of DSOR enzymes.

In the first step the NADPH binds to the fully oxidized enzyme. NADPH then reduces FAD to FADH2 and is followed by the nucleophilic attack of the C4A carbon of the FADH2 on the proximal cysteine thiol of the disulfide. In the next step, heterolytic cleavage takes place which leaves the FAD oxidized and the redox active disulfide in its fully reduced form. Although 2-KPCC belongs to the family of DSOR enzymes, there are key differences in the reaction catalyzed by 2-KPCC as compared to the other enzymes of the family. Firstly, 2-KPCC catalyzes a thioether bond cleavage rather than a disulfide bond. Secondly, after the thioether bond has been cleaved, the enzyme catalyzes a carboxylation reaction. A detailed description of the reactions catalyzed by 2-KPCC is shown in Figure16. The first half-reaction involves the transfer of reducing equivalents from flavin adenine dinucleotide to a cysteine that is a component of a disulfide bond. In the second half-reaction the substrate is reduced and the enzyme is regenerated. Steps 6a and 6b depict the fates of the α-carbanion formed in the reaction. It can be either protonated to form acetone when CO_2 is excluded from the enzyme or carboxylated to acetoacetate in the presence of CO_2. It is, however, advantageous to the organism to produce acetoacetate which can be easily shunted into the organism's central metabolism.

It is observed that in the presence of CO_2 there is very little or no acetone formed [38].

The crystal structure of 2-KPCC has been solved and provides insights into the mechanism of reductive cleavage of 2-ketopropyl coenzyme M. The overall structure of 2-KPCC is very similar to the dimeric structures observed for the classic members of the DSOR family [68] (Figure 17) namely glutathione reductase [69], trypanothione reductase [70] and lipoamide dehydrogenase [71]. Each monomer consists of three domains, including the FAD-binding domain, the NADPH binding domain, and the dimerization domain, or the interface domain.

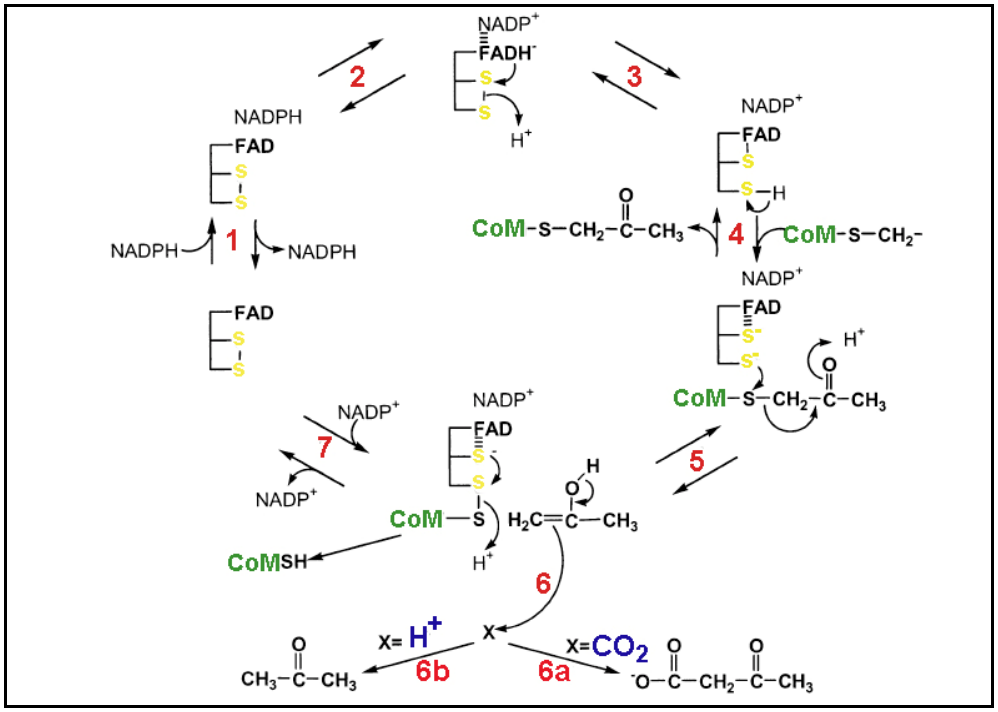

Figure 16. Proposed mechanism of action of 2-KPCC for the reductive cleavage and carboxylation of 2-ketoPropyl CoM.

The FAD and NADPH binding domain consists of standard Rossmann nucleotide binding motifs made up of large twisted β-sheets surrounded by α-helices. Although the overall structure of 2-KPCC is similar to the other enzymes of the DSOR family, there are some notable differences.

A 13-amino acid insertion can be found in 2-KPCC within the interface domain.

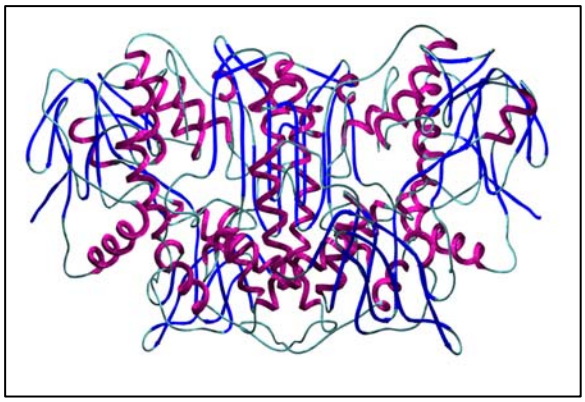

Figure 17. Overall tertiary structure of 2-KPCC.

This insertion loop is flanked by two cis-proline residues and eliminates the larger cleft that provides access to the active site for other DSOR enzymes. The substrate bound structure of 2-KPCC clearly shows that once the substrate is bound to the enzyme there is a conformational change which causes a collapse of the enzyme around the substrate which helps in eliminating the possibility of undue protonation of the substrate. In this way the enzyme also assures that there is a linear pathway for electron pair transfer between FAD, the redox active disulfide and the thioether bond of the substrate. The key interactions between the substrate and the enzyme are between the sulfonate group and two arginines.

The crystal structures of native and the substrate bound form of 2-KPCC reveal a mechanism which supports the proposed mechanism of the reaction shown in Figure16.

Carboxylation Reaction and Coenzyme M

Although epoxides are readily subjected to ubiquitous reactions like hydrolysis and glutathione conjugation, the alkene oxidizing bacteria such as *Xanthobacter* and *Rhodococcus* have chosen to carboxylate these compounds. There are a number of attractive features of epoxidation from a metabolic and toxicological stand point. The metabolic outcome of epoxide carboxylation is quite favorable and efficient as the end products like acetoacetate can be easily converted to central metabolites. If the epoxides are converted through reactions other than carboxylation there will be additional steps involved. For example, if epoxypropane is subjected to hydrolysis instead of carboxylation, in order to form succinyl-CoA, a central metabolite, the organism has to carry out subsequent steps of dehydration, oxidation, thiolation, carboxylation, mutation and isomerization. Therefore, it can be seen that in order to carry out epoxide metabolism through carboxylation the enzymes have favorably developed high affinities to their substrates. For example EaCoMT has a high affinity for epoxypropane as reflected by its low Km of 1.8 μM and its high levels of expression [30, 32].

Until 1999, only methanogenic archea were known to utilize coenzyme M [37]. It was first shown by Scott Ensign that coenzyme M is an important cofactor which is utilized in the carboxylation of epoxides in the bacteria namely *Xanthobacter* and *Rhodococcus* [30]. How these bacteria acquired the capability to synthesize coenzyme M or the original function of coenzyme M are questions which are unanswered. Made up of an ethyl spacer between sulfonate and thiol functional groups, coenzyme M is the smallest organic cofactor known. If the organism were to use hydrolysis reactions, there would be a need to handle bulky glutathione-alcohol conjugates formed by glutathione S-transferases. But with coenzyme M such problems would be circumvented because of the small 2-hydroxypropyl CoM conjugate formed in the reaction. The small size of 2-hydroxypropyl CoM is highly manageable and efficient for stereo selective dehydrogenation and reductive carboxylation in the subsequent steps. As seen in the substrate bound structure of 2-KPCC, the sulfonate moiety of CoM provides a selective handle for substrate recognition [66].

Metabolism of Secondary Alcohols and Ketones by Xanthobacter Autotrophicus

In general, acetone formed as a product of fermentative metabolism of certain anaerobic bacteria can accumulate to toxic levels if not metabolized further [14, 72, 73]. Acetone can have different fates in aerobic and anaerobic bacteria. In aerobes, acetone metabolism has been proposed to proceed by an O_2-dependent monoxygenase-catalyzed oxidation reaction, producing acetol as

the initial product. In other bacteria, including all anaerobes, acetone is carboxylated to acetoacetate or an acetoacetyl derivative [14, 72, 73]. *X. autotrophicus*, an obligate aerobe is capable of utilizing acetone or isopropanol as a sole source of carbon. The biological strategy for degradation of acetone/isopropanol is shown in Figure18. Isopropanol is converted to acetone by a secondary alcohol dehydrogenase and the resulting acetone is further metabolized [74, 75]. The secondary alcohol dehydrogenase of *X. autotrophicus* belongs to the family of medium chain dehydrogenases/reductases. Similar to SDR enzymes described earlier, MDR enzymes constitute a super family of enzymes with almost 1000 members reported [76]. Although most of the enzymes contain one catalytic zinc ion, there are some exceptions. All MDR enzymes are either NAD(H) or NADP(H) dependent. They are typically composed of ~350 amino acids and can exist as monomers, dimers or tetramers in solution. Each monomer is composed of two domains: an N-terminal nucleotide binding domain and a C-terminal substrate binding domain.

The enzyme carrying out the carboxylation of acetone to acetoacetate in *X. autotrophicus* is termed acetone carboxylase. Biochemical studies have suggested that the mechanism is analogous to the CO_2-dependent process occurring in anaerobic bacteria. Acetone carboxylase from *X. autotrophicus* strain Py2 is a heterohexameric protein comprised of three polypeptides with molecular weights 85,300, 78,300, and 19,600 daltons arranged in a $\alpha_2\beta_2\gamma_2$ quaternary structure. The preliminary biochemical and kinetic studies conducted

during initial characterization of acetone carboxylase provides direct evidence for involvement of ATP in acetone metabolism [14]. The carboxylation of acetone catalyzed by this enzyme is coupled to the obligate hydrolysis of ATP to 1 mol AMP and 2 mol of inorganic phosphate. Acetone carboxylation is a thermodynamically unfavorable process, thus the hydrolysis of ATP to ADP could provide the energy to drive carboxylation [77]. The observation that ATP is hydrolyzed to AMP and inorganic phosphate may suggest a dual role for ATP.

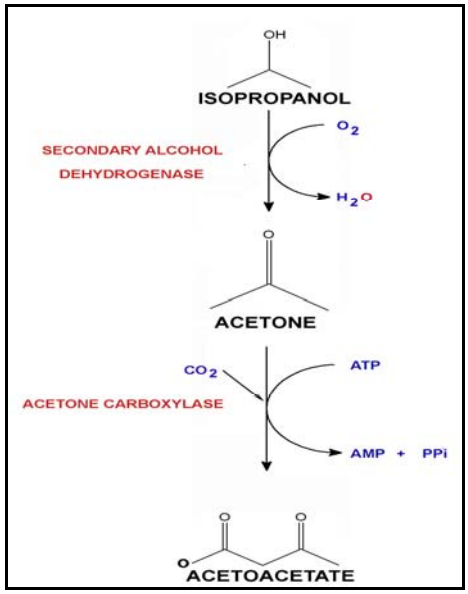

Figure 18. Metabolism of isopropanol and acetone in *Xanthobacter autotrophicus*.

It has been suggested that in addition to supplying the energy needs of the reaction, ATP could be involved in activating acetone for carboxylation via one or more group transfer reactions. This could occur through the transfer of a phosphoryl group to acetone. This process could potentially be mediated through a phosphoryl-enzyme intermediate. Since acetone carboxylation likely involves

the attack of the carbanion of acetone on CO_2, the reaction could be facilitated by the stabilization of the enol tautamer of acetone through phosphoryl-group transfer to the oxygen atom of the enolate. This proposed reaction scheme has similarities to that of the reaction scheme of the well characterized phosphoenolpyruvate synthase which catalyzes the conversion of pyruvate to phosphoenol pyruvate with the concomitant hydrolysis of ATP to AMP and inorganic phosphate [17].

Research Directions

The broad, long term goal of this research is to provide insight into the mechanism of novel carboxylation reactions and chiral discrimination that occurs in the enzyme involved in bacterial alkene, ketone and epoxide metabolism. Aliphatic alkenes as well as aromatic compounds are substrates for microsomal cytochrome P450 oxygenases. These compounds are oxygenated across double bonds to the corresponding epoxides. The epoxides formed serve as alkylating reagents that can covalently modify both proteins and nucleic acids and result in mutagenic effects. Despite their mutagenic effects, some bacteria like *X. autotrophicus* and *R. rhodochrous* are capable of using epoxides and ketones as sole sources of carbon. Studies on the novel bacterial pathway of metabolism of alkenes and ketones in *X. autotrophicus* and *R. rhodochrous* have identified a central role for CO_2 and specifically CO_2 fixation reactions in these processes. The CO_2 fixing enzymes of the pathway are distinct carboxylases with unique

molecular properties and cofactor requirements. Although there is a wealth of biochemical and kinetic data available for these enzymes, much less is known about the mechanism of stereo chiral discrimination and also the terminal carboxylation reactions involving 2-KPCC and acetone carboxylase. Structural studies of these enzymes will provide in depth understanding of the catalytic mechanism.

The research reported in this dissertation focuses on providing the structural basis for the mechanistic understanding of the stereo specific catalysis and novel carboxylase reactions in the convergent pathways of alkene and ketone metabolism in *X. autotrophicus*. X-ray crystallography has always been a powerful tool in providing insights into the catalytic mechanism of enzymes. Determination of the structures of these enzymes is the key not only in confirming kinetic and biochemical studies but also in revealing novel structural features that confer unique catalytic properties to the enzymes. The determination of the structure of the *R*- and *S*- dehydrogenases could shed light on the stereospecific catalysis and the mechanism of chiral discrimination in these enzymes. These enzymes share a high degree of sequence identity, and key differences are present in the C-terminal region of the protein. It would be interesting to see how two enzymes having very few differences in the protein sequence can discriminate structurally related substrates and bring about stereo specific dehydrogenation.

Since the unusual methanogenic cofactor CoM is utilized in the epoxidation carboxylation pathway, it would be interesting to see if there are structural features in the enzymes involved in the pathway which allow them to utilize CoM. A detailed analysis of the structures with respect to CoM binding could provide such an insight. The culminating reactions of both propylene and acetone metabolism in *X. autotrophicus* are novel carboxylation reactions involving the enzymes 2-KPCC and acetone carboxylase, respectively. The structure of 2-KPCC has already been solved and has made significant contributions towards the elucidation of the detailed reaction mechanism. While the structure of 2-KPCC has been captured in many states including the ones with substrate and analogs bound to the enzymes, structural studies of acetone carboxylase have lagged behind. The determination of the structure of acetone carboxylase will provide significant details regarding the less known mechanism of this enzyme.

CHAPTER 2

STRUCTURAL BASIS FOR STEREOSLECTIVITY IN THE (*R*) – AND (*S*) - HYDROXYPROPYLTHIOETHNAESULFONATE DEHYDROGENASES

Chapter Abstract

Epoxide metabolism in *Xanthobacter autotrophicus* Py2 results in the conversion of epoxypropane to acetoacetate. Epoxide metabolism is initiated by the nucleophilic addition of coenzyme M to the *(R)*- and *(S)*-enantiomers of epoxypropane to form the respective enantiomers of 2-hydroxypropyl-coenzyme M. The (*R)* and (*S)* enantiomers of 2-hydroxypropyl coenzyme are oxidized to the achiral product 2-ketopropyl-CoM by two stereoselective dehydrogenases. The dehydrogenases catalyzing these reactions, termed *R*-hydroxypropyl coenzyme M dehydrogenase (*R*- HPCDH) and *S*- hydroxypropyl coenzyme M dehydrogenase (*S*-HPCDH) are NAD^+-dependent enzymes belonging to the short-chain dehydrogenase/reductase (SDR) family of enzymes. In the present study, the crystal structure of *R*- HPCDH co-crystallized in the presence of *S*-hydroxypropyl coenzyme M has been determined using X-ray diffraction methods and refined to 1.8 Å resolution. The structure of *R*-HPCDH is tetrameric and stabilized by the interaction of the terminal carboxylates of each subunit with divalent metal ions. The structure of the presumed product bound state reveals that binding interactions between the negatively charged oxygen atoms of the sulfonate moiety have striking similarities to sulfonate interactions

observed in the previously determined structure of 2-ketopropyl-CoM oxidoreductase / carboxylase, highlighting the utility of coenzyme M as a carrier molecule in the pathway. The key elements of the aforementioned interactions are electrostatic interactions between the sulfonate oxygen atoms and two arginine residues (R152 and R196) of *R*- HPCDH. The comparison of the structure of *R*- HPCDH with a homology model of the *S*-HPCDH provides a structural basis for a mechanism of substrate specificity in which the binding of the substrate sulfonate moiety at distinct sites on each stereoselective enzyme directs the orientation of the appropriate substrate enantiomer for hydride abstraction.

Introduction

The highly reactive nature of epoxides makes them valuable intermediates for the synthesis of more complex optically active or biologically active compounds, including pharmaceuticals and agrichemicals. For this reason there has been considerable interest in exploiting epoxide utilizing enzymes which have stereoselective and regioselective properties for producing enantiopure compounds. Some of the known alkene oxidizing enzymes, such as heme-dependent monooxygenases, ω-hydroxylases, and non-heme oxygenases, catalyze direct stereospecific epoxidation of alkenes [78]. Alternatively, enantiopure epoxides can be resolved from racemic mixtures by enantioselective epoxide metabolizing enzymes. Some microorganisms are capable of utilizing aliphatic

alkenes and epoxides as sole carbon sources [10]. The pathway for the metabolism of short chain alkenes and epoxides in aerobic bacterium *Xanthobacter autotrophicus* Py2 has been well characterized [29-34]. Integral to this metabolic pathway are the two stereospecific 2-Hydroxypropyl thioethanesulfonate dehydrogenases designated *R*-HPCDH and *S*-HPCDH that individually allow for the oxidation of only *R* or *S* epoxypropane enantiomers, respectively to the common achiral product 2-ketopropyl thioethanesulfonate or 2-ketopropyl CoM.

The two dehydrogenases *R*-HPCDH and *S*-HPCDH are related by 41% sequence similarity and belong to the short chain dehydrogenases reductases super family of enzymes [29]. Enzymes of this class are approximately 250 amino acids in length and exist as dimers or tetramers in solution [47]. The culmination of structural, kinetic and mechanistic studies on the SDR enzymes has resulted in the formulation of a general mechanism of action presented in Figure 19 [52]. Based on this and sequence alignments along with site directed mutagenesis studies, the identity and function of the catalytic residues have been established for the *R*-HPCDH [51]. The residues Tyr155, Lys159 and Ser142 form the catalytic triad and their functions are consistent with the general mechanism (Figure 19). According to this, the tyrosyl hydroxyl is deprotonated which then serves as a general base for proton abstraction from the hydroxyl group of the substrate. The amino group of lysine stabilizes the deprotonated

Tyr general base. Serine is thought to increase the acidity of the substrate hydroxyl through hydrogen bonding.

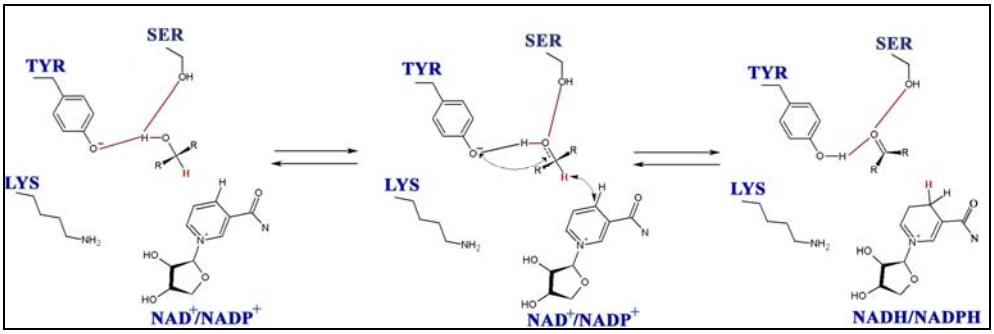

Figure 19. General mechanism of SDR enzymes involving the conserved serine, tyrosine and lysine catalytic triad.

Although the dehydrogenases are in most cases stereospecific, there is only one other case known in which pairs of stereospecific dehydrogenases act in concert in one pathway. Tropinone reductase -I and II from *Datura stramonium* (TR-I and TR-II) catalyze the reversible stereospecific reduction of the 3-keto group of tropinone to form the enantiomers tropine (α-hydroxy) and pseudotropine (β-hydroxy), respectively [61, 62]. Structural characterization of TR-I and TR-II has revealed that electrostatic interactions between substrates and charged amino acids in the active site were responsible for orienting the substrate to its correct geometry. For TR-I, the positively charged histidine nitrogen (His112) repels the positive charge on the nitrogen atom of tropinone and serves to orient the substrate in the active site. In the case of TR-II, the negatively charged glutamate carboxylate (Glu-156) attracts the substrate nitrogen and thus fixes its position.

Our recent work involving amino acid modification in the *R*-HPCDH enzyme has implicated the involvement of two arginine residues (Arg-196 and Arg152) in substrate binding [67]. These residues have been proposed to support substrate binding through electrostatic interactions between the positively charged arginine side chains and the negatively charged oxygen atoms of the substrate sulfonate group. In the context of this proposal, our structural studies have revealed that the binding of substrate for 2-KPCC, which catalyzes the next step in the pathway, is mediated by interactions between two arginine residues (Arg-56 and Arg-365) and the substrate sulfonate oxygens [66]. The negatively charged sulfonate group offers a convenient "handle" on the substrate molecule for binding to the active site. Presumably, similar interactions between the sulfonate binding group and positively charged amino acid side chains are involved in the binding of substrate in both *R*- and *S*-HPCDH. Based on the aforementioned site directed mutagenesis studies of *R*-HPCDH and the crystal structure of 2-KPCC, a model has been proposed for the basis of stereoselectivity in which distinct sulfonate binding sites in the *R*- and *S*-HPCDH promotes the correct orientation for hydride abstraction of their respective substrates (Figure20).

For the tropinone reductases, the three dimensional structures were the key to defining the mechanism and structural basis of stereoselectivity. In the present study, the three dimensional structure of *R*-HPCDH reveals the features of the substrate binding site that contribute to stereoselective hydride

abstraction. A combination of substrate docking and homology modeling studies together with the structural analysis of the *R*-HPCDH indicates that the difference in binding of the negatively charged sulfonate group of the substrate in the C-terminal substrate binding domain of each enzyme is the key to chiral discrimination.

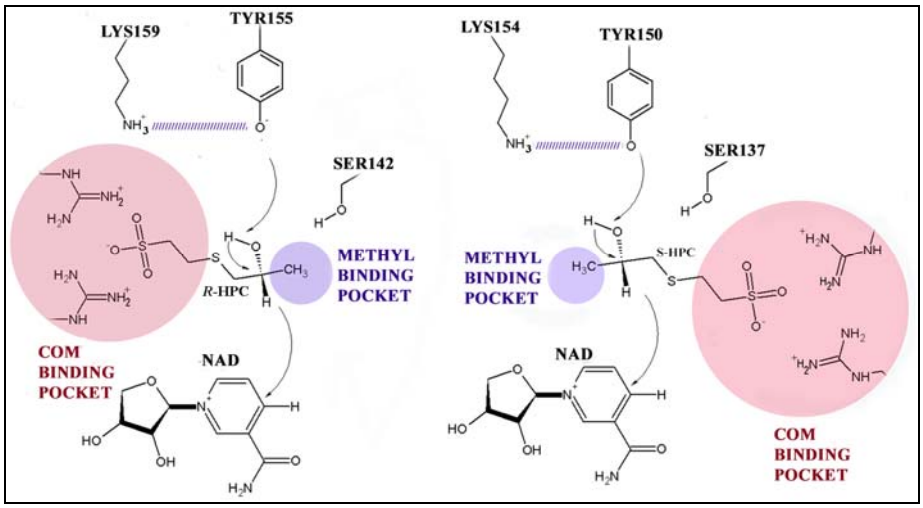

Figure 20. Models for substrate binding to (A) *R*-HPCDH and (B) *S*-HPCDH for stereo specific dehydrogenation of the enantiomers of 2-hydroxypropyl CoM.

Experimental Procedures

Crystals of 2-*R*-HPCDH can be obtained in three crystal forms, two of which exist with $P2_1$ space group symmetry and one with *C*2 [79]. The *C*2 crystal form grown in the absence of substrates or inhibitors is of poor to moderate diffraction quality such that data to ~3 Å resolution can be collected routinely at a synchrotron source. A $P2_1$ crystal form of similar quality can also be grown in the absence of substrates or inhibitors. An additional $P2_1$ crystal

form of significantly better diffraction (better than 2.0 Å resolution) is obtained when crystals are grown in the presence of substrates or inhibitors and NAD^+ [79]. Although the primary sequence is considerably identical with the sequences of members of the SDR family, attempts to determine the structure by molecular replacement methods in any of the aforementioned crystal forms were unsuccessful. A significant impediment to solving the structure by molecular replacement was most likely that each crystal exists with multiple (4-8) SDR subunits per asymmetric unit. Since molecular replacement did not yield a reliable solution, the structure was determined by the multiple wavelength anomalous diffraction (MAD) methods. MAD studies were conducted using a mercury heavy atom derivative of the enzyme. To generate the *R*-HPCDH mercury derivative, native $P2_1$ crystals were grown in the presence of 2mM EMTS with Hampton cryo condition 9 [80]. A three wavelength MAD data set was collected at SSRL beam line 9-2 equipped with a mar345 imaging plate detector (Mar USA, Inc.). Data were processed with DENZO [81] and scaled with SCALPACK [81] of the HKL or HKL2000 [81] program suite and the data collection statistics are summarized in Table 1.

The positions of heavy atom binding sites in the MAD experiment were identified using SOLVE [82]. Initial searches and refinement of nine heavy atom sites per unit cell resulted in a figure of merit of 0.46 from 20.0 Å to 3.2 Å resolution. The heavy atom sites were utilized in RESOLVE [82] to identify non crystallographic 2-fold operation and resulted in an improvement of figure of

merit of 0.60. The automated model building routine of RESOLVE could produce a partial model of the structure and the remaining regions of the molecule were built manually into the initial phase extended maps using the model building program O [83]. The structure refinement was carried out using CNS [84] from 50 to 3.0 Å with a 0σ cutoff with 5% of the data set aside for the calculation of R_{free} [85]. The structure was subjected to multiple rounds of refinement and model building in order to generate a model suitable for a molecular replacement search model for the other crystal forms.

Table 1. Multiple wavelength anomalous diffraction data statistics.

Cell dimensions	a = 67.149 Å		
	b = 124.178 Å		
	c = 121.908 Å		
	β = 100.742°		
	Peak	Remote	Inflection
	λ=0.87008 Å	λ= 1.00871 Å	λ=0.87278 Å
Resolution (Å)	50.0-3.00	50.0-3.00	50.0-3.00
Completeness (%)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)
Observed reflections	772,150	921,887	1,440,602
Unique reflections	290461	276477	264,138
I/σ	8.5 (8.4)	9.7 (9.1)	10.1 (6.5)
R_{merge}[a]	0.066 (0.167)	0.072 (0.176)	0.077 (0.184)

Numbers in parentheses indicate values for the highest resolution bin.
a. $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle|/\Sigma_{hkl}\Sigma_i|\langle I\rangle|$, where I_i is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.

When the R_{free} of the native $P2_1$ crystal structure dropped just below 30% this structure was used as a search model for molecular replacement to solve the structure of the substrate and inhibitor bound enzyme. Using CNS, a solution with four subunits in the asymmetric unit resulted in a correlation coefficient of 0.659 and an *R* factor of 0.381 with very good protein packing interactions. A well-determined structure was obtained through multiple rounds of model building using the program O followed by refinement using CNS for data from 20 Å to 1.8 Å with a 0σ cutoff. The data statistics and the refinement statistics of the native and the substrate/product co-crystal structure are summarized in Tables 2 and 3, respectively. The final models obey reasonable stereochemistry with 100% of residues occupying allowed regions of a Ramachandran Plot calculated using PROCHECK [86]. The final structure of the *R*-HPCDH together with similar structures like mannitol dehydrogenase and levodione reductase [87] were then used as templates to generate a homology model of the *S*-HPCDH using the CPHMODEL [88] and the SWISS MODEL servers [89]. All figures were generated using SWISS PDB VIEWER [90] rendered using POVRAY [91].

Table 2. Data statistics for the *R*-HPCDH Co-crystallized with *S*-HPC.

Space group	$P2_1$
Cell dimensions	a = 64.44 Å b = 110.28 Å c = 68.98 Å β = 93.88°
Resolution (Å)	20.00-1.8

Table 2. Data statistics for the *R*-HPCDH Co-crystallized with *S*-HPC-continued.

Completeness (%)	93.9 (95.2)
Observed reflections	149,113
Unique reflections	71,201
I/σ	6.5 (2.1)
R_{merge} [b]	0.084 (0.303)

Numbers in parentheses indicate values for the highest resolution bin. b. $R_{merge} = \Sigma hkl \Sigma i |Ii - \langle I \rangle| / \Sigma hkl \Sigma i |\langle I \rangle|$,
where Ii is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.

Table 3. Refinement statistics for the *R*-HPCDH Co-crystallized with *S*-HPC.

Resolution (Å)	20.00-1.80
R_{free}	0.227
R_{cryst}	0.197
No. Non hydrogen atoms	
Protein	4004
Cofactor	76
Product	44
Solvent	915
RMSD from target values	
Bond lengths (Å)	0.012
Bond angles (°)	1.55
Average isotropic B factor	
All atoms	21.83
Protein main chain	16.94
Protein side chain	21.58
NAD^+	55.98
KPC	74.22
Water	35.08

Results and Discussions

Overall Structure

The tertiary structure of each *R*-HPCDH subunit is a classic Rossmann fold common for the SDR family of enzymes. It has a α/β type structure with an open twisted parallel 7-stranded β sheet in the middle surrounded by 6 α helices (Figure 21A). Each subunit consists of two βαβαβ motifs, βA-αB-βB-αC-βC and βD-αE-βE-αF-βF, exhibiting right-handed crossover [92] by a seventh β strand with a left-handed cross over connection between β strands F and G (Figure 21 B). The active site lies in the central region of each subunit and the catalytic residues Ser142, Tyr155 and Lys159 are located on βE and αE. The active site is located in a deep cleft which is lined on one side by two short helices αFF1 (residues 192 to 201) and αFF2 (residues 202 to 211) (Figure 21 C). The analogous region in other SDR enzymes has been shown to be responsible for substrate specificity [93]. The other side of the active site cavity is lined by the loop formed between βE and βF (Residues 143 to 153). The residues of the active site are very well resolved and the electron density map in the region is excellent allowing unambiguous assignment of positions of all the atoms in the side chains (Figure 21 D). The position of the active site residues is consistent with their function in the proposed general mechanism of the SDR enzymes (Figure 19). Tyrosine acts as a general base for proton abstraction from the alcohol substrate while lysine159 is important in cofactor binding and in

lowering of the pKa of the hydroxyl group of the tyrosine. The serine142 coordinates the substrate hydroxyl group.

The quaternary structure of *R*-HPCDH (Figure 22 A) is very similar to the tetrameric structures observed for other members of the SDR family including mannitol dehydrogenase (MDH) [60] (Figure 22 B) and haloalcohol dehydrogenase (HADH) [94]. Previously, size exclusion chromatography suggested that *R*-HPCDH exists as a dimer in solution [29] however, the crystal structure suggests that enzyme exists as a tetramer. The three dimensional structure of *R*-HPCDH indicates four subunits related by 222 symmetry with dimensions of 71 Å x 61 Å x 59 Å. Interactions between individual monomers are along three non-crystallographic two fold axes analogous to that of the previously characterized SDR tetramers [60, 94].

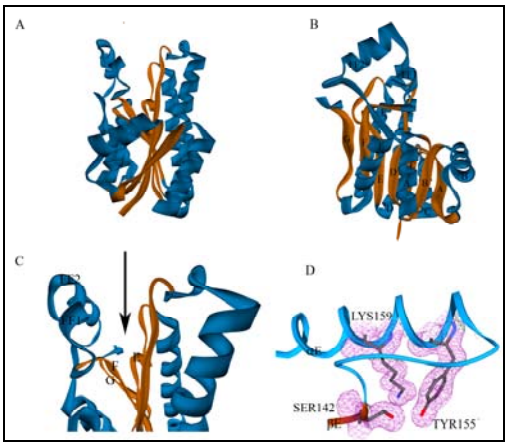

Figure 21. (A) Overall tertiary structure of one of the subunits of *R*-HPCDH with β-sheets colored in cyan and α-helices in cyan displaying the Twisted beta sheet or the typical Rossmann fold. Bound NAD^+ is shown in Corey-Pauling-Koltun (CPK) representation. (B) Alternate view of the same subunit. (C) Zoomed in view of the same subunit with arrow pointing towards the substrate binding cleft. (D) Electron density maps (2*F*o-*F*c) contoured at 1σ about the residues forming the catalytic triad of the *R*-HPCDH.

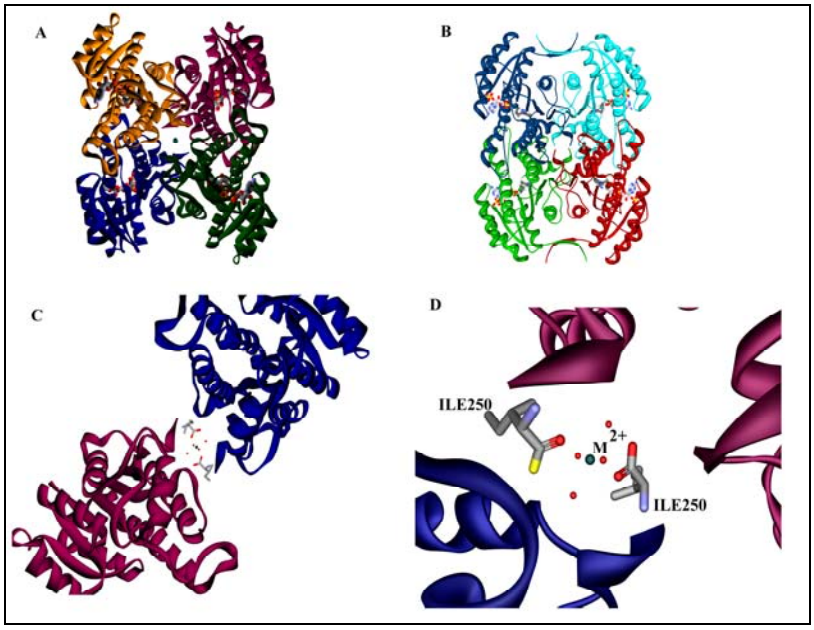

Figure 22. (A) Overall fold of the tetrameric *R*-HPCDH with individual subunits colored in blue, green, orange, and magenta. (B) Overall fold of the well-characterized mannitol dehydrogenase in the same orientation as the structure of *R*-HPCDH shown in panel A with the individual subunits shown in blue, cyan, green, and rust. (C) Two subunits of *R*-HPCDH held together at the C-terminus by a divalent metal shown in blue via the terminal isoleucine residues. (D) Zoomed in view of the divalent metal site coordinated by carboxylates of isoleucines supplied by two subunits of the enzyme and four water molecules at the equatorial position.

The degree of interaction between the subunits of the putative tetramer is far too extensive for typical crystal packing interactions and involves a number of hydrophobic interactions. The specific interactions that stabilize the tetramers are very similar in MDH, HADH, and *R*-HPCDH. Interestingly, the tetrameric structure appears to be stabilized by the interaction of the C-termini with what are presumed to be divalent metal ions (Figure 22C). The metal ions are coordinated in a monodentate fashion at the axial position by one of the oxygens of the C-terminal carboxylate group of Ile250 supplied by each monomer and by

four water molecules at the equatorial positions (Figure 22 D). An analogous metal coordination is observed in the structure of mannitol dehydrogenase where the C-terminal tryptophans were observed to be coordinated by nickel ions [60]. The identity of the metal ion in *R*-HPCDH has not been determined.

Interactions of NAD^+ with Protein

The interactions of NAD^+ with the enzyme are summarized in Figure 23. The NAD^+ binding domain consists of a standard Rossmann nucleotide binding motif of a large twisted β sheet surrounded by α helices. NAD^+ is bound in an extended conformation at the C terminal end of the β sheet. The electron density for the adenine ring and its associated ribose is of high quality permitting unambiguous identification of the positions of these portions of the cofactor (Figure23 A). Poor maps have resulted in weaker electron densities around the nicotinamide ring and its associated ribose ring. Comparison of the conformation and position of the adenine moiety $NADP^+$ bound to MDH with the adenine of NAD^+ bound structure of *R*-HPCDH are very similar. Although density was rather poor for the nicotinamide in the *R*-HPCDH structure, being able to position the adenine of NAD^+ with a high degree of confidence places limited degrees of freedom in the positioning of nicotinamide in the structure. The limited degrees of freedom in combination with the anticipated positioning of nicotinamide with respect to the catalytic triad which is conserved for all SDR enzymes (Ex. MDH) allows a relatively unequivocal assignment of the position of NAD^+.

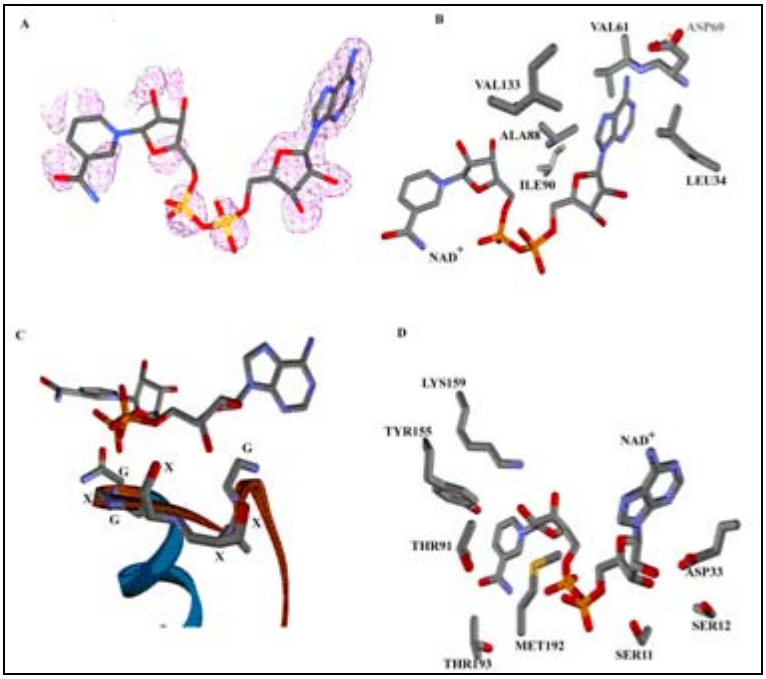

Figure 23. (A) 2*F*o-*F*c simulated annealing omit electron density map around the NAD^+ molecule contoured at 1σ show unambiguous density around the adenine, first ribose and phosphate and weak density for nicotinamide ring and the second ribose. (B) The hydrophobic pocket in which the adenine ring of the NAD^+ resides is shown. The side chains of the amino acids involved in interactions are displayed. (C) The classic GXXXGXG Glycine-rich NAD^+ binding motif in *R*-HPCDH with NAD^+ bound. (D) View of the residues interacting with the nicotinamide ring of the NAD^+.

The adenine ring of NAD^+ binds in a hydrophobic pocket on the enzyme surface formed by the side chains of Ile90, Val113, Val61, Leu34, Ala88 and Lys109 (Figure 23 B). Hydrogen bonds are formed by adenine N1 and Asp60 side chain and N6 and peptide nitrogen of Val61. The sequence Gly9XXXGly13XGly15 corresponding to the classic glycine-rich consensus sequence GlyXXXGlyXGly present in $NAD(P)^+$ binding dehydrogenases can be seen interacting with the adenine ribose of the NAD^+ (Figure 23 C). Also the 2' and 3' hydroxyl groups of the adenine ribose form hydrogen bonds with an aspartate (Asp33) carboxylate

group as is common for SDR enzymes selectively binding NAD^+ and not $NADP^+$ [55]. The enzymes utilizing $NADP^+$ have a positively charged residue such as arginine at that position to interact with the negatively charged phosphate. The 2' and 3' hydroxyl groups of the ribose also interact with Ser12 via hydrogen bonds and indirectly with Ser11 through a water molecule. A water molecule at this position is a structural feature of all Rossmann dinucleotide-binding domains [95]. The phosphate groups form hydrogen bonding interactions with side chains of Ser12, Pro191, Met192, Thr190 and Asn14. The 2' and 3' hydroxyls of the second ribose (attached to the nicotinamide ring) form a bifurcate hydrogen bond with Lys159. Additional interactions that stabilize the NAD^+ in the enzyme include hydrogen bonding of the active site tyrosine with the 3' hydroxyl of the second ribose and hydrogen bonding between Thr190 and Met192 and the amide group of nicotinamide (Figure 23 D).

Substrate Binding Site

Since crystals were grown aerobically in the presence of the substrates NAD^+ and *R*-HPC, it is anticipated that the substrate *R*-HPC would have been largely oxidized to the product 2-KPC during the course of the crystallization. The quality of the electron density maps for the sulfonate and the thioether portion of 2-KPC is excellent, but the ketopropyl moiety of the product is not well resolved (Figure 24 A). The product, 2-KPC is bound to the enzyme in the aforementioned substrate binding cleft formed by the loop between βF and βG and the loop between βE and βF. This cleft lies directly above the B face of the

nicotinamide ring and the substrate binding site is mainly surrounded by hydrophobic residues. Specific interactions are mainly between two Arg residues and the product sulfonate. The binding of the product 2-KPC to *R*-HPCDH (Figure 24 B)is nearly analogous to the binding of substrate 2-KPC-to 2-KPCC [66] (Figure 24 C). The substrate binding site of the *R*-HPCDH is characterized by two methionines (Met 192 and Met 187) which flank the CoM moiety of the substrate. The sulfonate is bound to two arginine side chains, Arg152 and Arg196, by salt bridge interactions with two methionine residues (Met192 and Met187) that appear to act in shielding the sulfonate charge from the active site. Recent site directed mutagenesis studies have implicated these arginine residues in sulfonate binding [67]. These arginines are located on the loops between βE and αE and βF and αF, respectively and approach the substrate/product from the sides. Trp195 appears to form a 'backstop' that may prevent translation of the substrate and facilitate proper alignment of the substrate with respect to NAD^+ for direct hydride transfer. Comparison of the substrate binding site in both these enzymes suggest a general architecture of flanking methionines and positively charged amino acids directly binding the sulfonate moiety might be useful in identifying sulfonate binding site in other CoM enzymes. To envision how the enzyme binds the substrate, we manually modeled *R*-HPC at the active site by superimposing the sulfonate moiety and the thio ether group of the substrate on those of the product. Orientation of the hydroxyl group of the substrate towards the hydroxyl of the Ser142 and Tyr155

resulted in the correct orientation of the hydrogen atom of the substrate towards the NAD^+ B-face for direct hydride transfer (Figure 25).

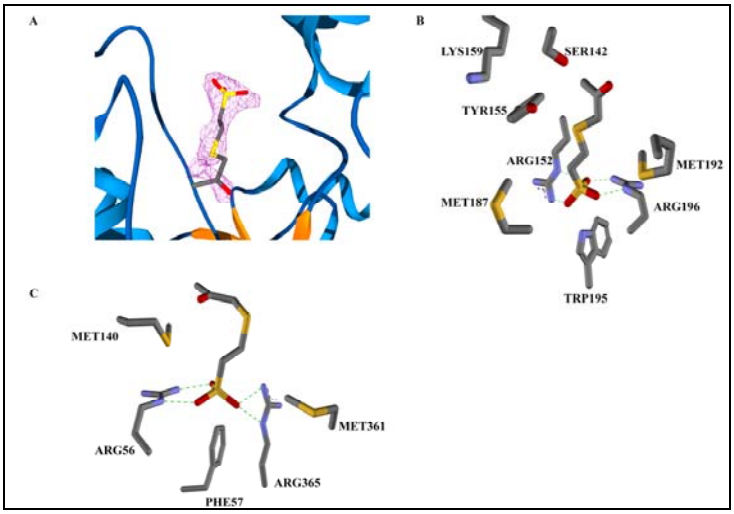

Figure 24. (A) Overall view of the product bound in the substrate binding cleft with a 2*F*o-*F*c simulated annealing omit electron density map contoured at 1σ about the product. (B) Product binding at the active site of *R*-HPCDH highlighting the specific interactions with Arg side chains and adjacent tryptophan residue. The methionines flanking the product are also shown.(C) Substrate binding at the active site of 2-KPCC showing the Arginines and the flanking methionines.

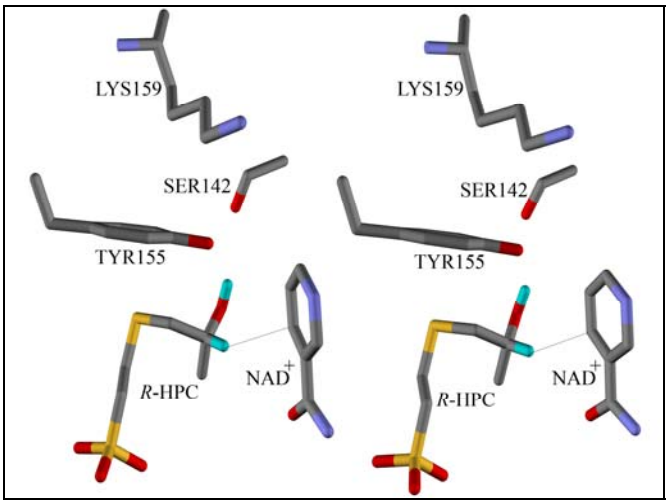

Figure 25. Wall-eyed stereo drawing of *R*-HPC modeled at the active site of *R*-HPCDH showing active site residues and nicotinamide ring of the NAD^+ molecule.

Kinetic analysis shows that the *S*-HPC is a very strong competitive inhibitor of *R*-HPCDH with K_i very similar to the K_m of the *R*-HPC [29]. Although the *S*-HPC binds the *R*-HPCDH with high affinity, the activity of the *R*-HPCDH with the *S*-HPC is only 0.1 % compared to its activity with the *R*-HPC. In order to visualize the binding mode of *S*-HPC at the active site of the *R*-HPCDH, we manually modeled *S*-HPC at the active site by superimposing the *S*-HPC on *R*-HPC. Positioning of the hydroxyl group of the *S*-HPC exactly over that of *R*-HPC orients the hydrogen atom of the *S*-HPC away from the NAD^+ providing the structural basis for *S*-HPC as a competitive inhibitor (Figure 26).

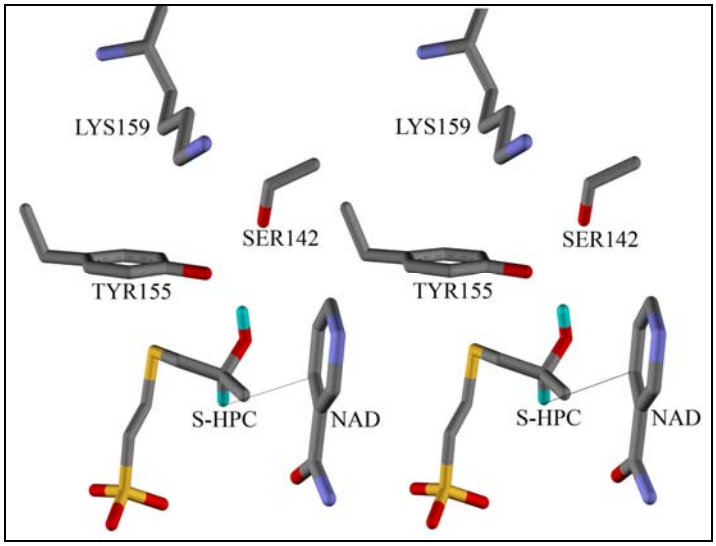

Figure 26. Wall-eyed stereo drawing of the *S*-HPC bound at the active site of *R*-HPCDH in the same orientation as shown in panel A. Nicotinamide part of the NAD^+ is shown in both panels.

Structural Basis for Stereos Electivity

One of the main goals of our structural analysis of the *R*- and *S*-HPCDH enzymes is to reveal the mechanism of chiral discrimination for these enzymes.

The specific interactions observed between 2-KPC and *R*-HPCDH in the structure described herein as well as previous biochemical studies suggest that substrate sulfonate binding is key to orienting the substrate at the active site for hydride abstraction. The structures of both *R*-HPCDH and 2-KPCC illustrate the effectiveness of CoM as a cofactor in this biochemical pathway. If indeed sulfonate binding is the key to correct substrate alignment, then one would expect that sulfonate binding and the specific sulfonate binding sites would be different in *R*-HPCDH and *S*-HPCDH. In fact, the idea that chiral discrimination is conferred through distinct substrate binding sites had been previously proposed from the results of our biochemical studies and amino acid sequence comparisons [29].

Since a direct crystallographic analysis of *S*-HPCDH has not been possible, the three dimensional structures of *R*-HPCDH and other SDR enzymes were used to construct a homology model of the *S*-HPCDH (Figure 27 A). Superimposition of *S*-HPCDH on *R*-HPCDH reveals differences at the substrate binding site that highly support sulfonate binding as the key to stereoselectivity (Figure 27 B). Although residues in the active site are superimposable (Figure 27 C), the Arg residues (Arg152 and Arg196) that were observed to be key in binding the sulfonate in *R*-HPCDH have been replaced in *S*-HPCDH by Met147 and Gln191, respectively (Figure 27 D). These side chains would not support the binding of substrate in the same manner as observed for *R*-HPCDH.

The absence of the sulfonate binding site at that region clearly suggests that the *S*-HPC substrate cannot be bound in the similar binding mode as the *R*-HPC bound to the *R*-HPCDH. Effective hydride abstraction by short chain dehydrogenases is dependent on the correct orientation of the hydride with respect to the nicotinamide and the correct orientation of the substrate hydroxyl group with respect to the Tyr of the catalytic triad. With this in mind we have utilized the homology model of the *S*-HPCDH and the structure of the *S*-HPC substrate to attempt to implicate a sulfonate binding site in this enzyme.

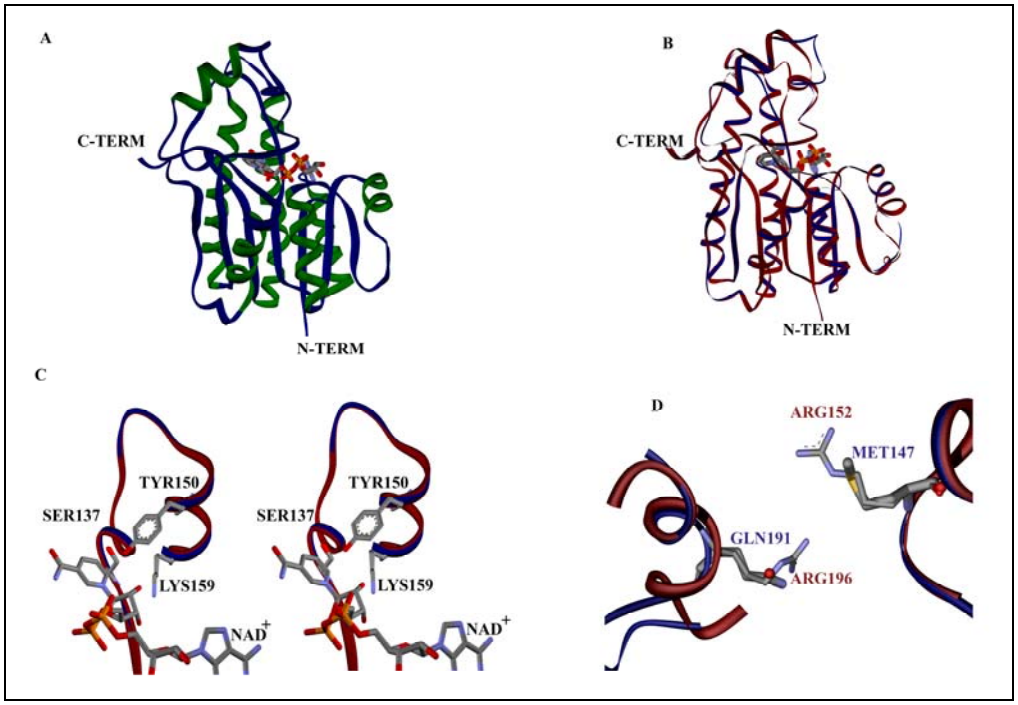

Figure 27. (A) Overall tertiary structure of *S*-HPCDH with β-sheets colored in blue and α-helices in green. (B) Superimposition of the structure of *S*-HPCDH (blue) on *R*-HPCDH (dark pink) to show differences in the substrate binding region. (C) View of the active sites of superimposed *R* and *S*-HPCDH showing similar orientation of NAD^+ in both enzymes. (D) View of *S*-HPCDH superimposed on *R*-HPCDH showing differences in the amino acid residues in the sulfonate binding region. Superimposition was accomplished as described by Kabsch [Kabsch, 1976 #32].

In order to achieve the proper position of the substrate hydroxyl and hydrogen, the spatial orientation of the other two groups, namely the sulfonate and methyl group, must be different in the two enzymes. Since it is hypothesized that differential placement of positively charged residues within the substrate binding region of the enzyme provides the "switch" required for chiral discrimination, the structure of *S*-HPCDH was examined for the presence of a putative sulfonate binding site. Superimposition of *S*-HPC bound to *R*-HPCDH on the NAD^+ bound structure of *S*-HPCDH followed by rigid body reorientation so that the hydroxyl and hydrogen groups with the active site residues and NAD^+ puts the sulfonate moiety within hydrogen bonding distance to two positively charged residues (Arg205 and Lys208) (Figure 28).

These residues reside on one side of the substrate binding cleft rather than approaching the sulfonate groups from both sides as in case of the *R*-HPCDH. As in the case of both *R*-HPCDH and 2-KPCC, again substrate binding is characterized by two positively charged sulfonate binding groups shielded by two flanking methionine residues (Met147 and Met148). The sulfonate binding sites in the *R*- and *S*-HPCDH represent analogous sites involving a different set of amino acid residues at spatially different positions on the respective enzymes. We were able to compare the spatial orientation of this putative sulfonate binding site to the sulfonate binding site of *R*-HPCDH to envision the differential placement of positively charged residues that bind the sulfonate moiety of the substrate by superimposing the *S*- and *R*-HPCDH enzymes. Superimposition of

the *R* and *S*–HPCDH clearly indicates the differential spatial orientation of the sulfonate binding sites in these two enzymes. From the Figure 29, it is clear that although the orientation of active site residues and the NAD^+ are exactly same, the sulfonate and the methyl groups are switched in order to position the hydrogen and hydroxyl groups of the substrate for hydride and proton abstraction, respectively.

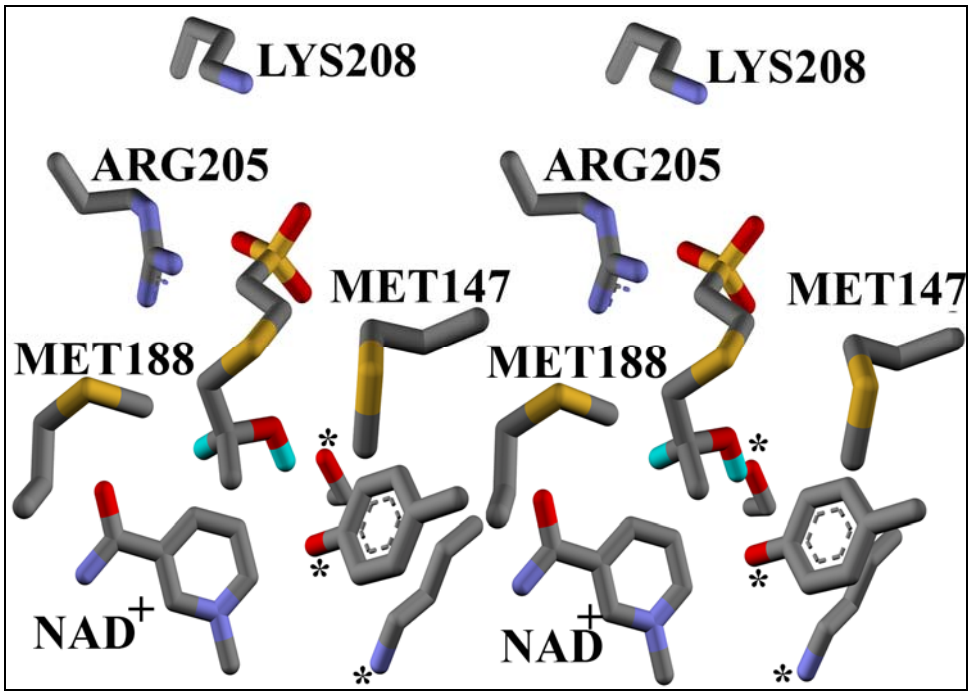

Figure 28. Wall-eyed stereo picture of the putative substrate binding site in *S*-HPCDH. The Serine-Tyrosine-Lysine catalytic triad is marked by asterisks. Nicotinamide part of the NAD^+ is shown.

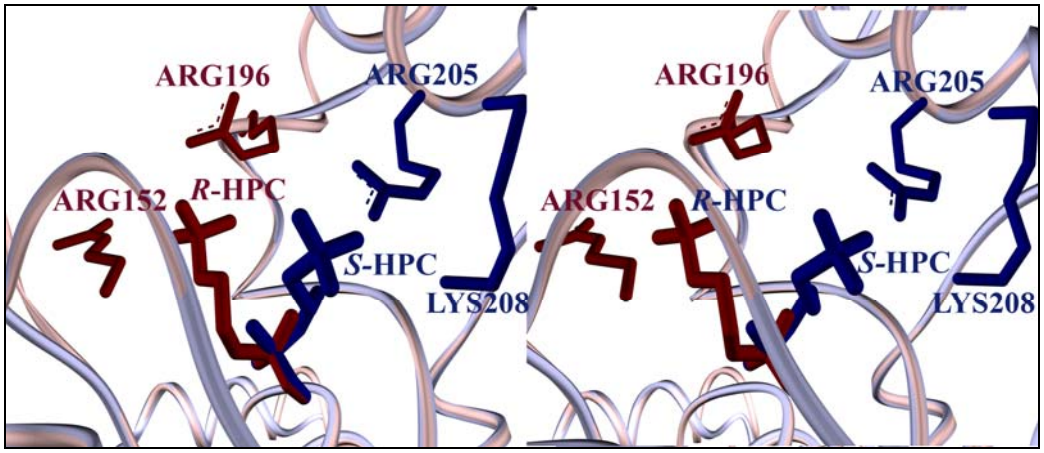

Figure 29. Wall eyed stereo picture of superimposed *R* and S-HPCDH to show the differential spatial orientation of sulfonate binding sites. The *R*-HPCDH residues and *R*-HPC are colored dark pink while the *S*-HPCDH residues and *S*-HPC are colored blue.

The results discussed herein provide a structural basis for chiral discrimination of the *R*- and the *S*-HPC substrates by the stereo selective dehydrogenase pair and provide support for our previously proposed model [67]. Although the overall structure and architecture of both the enzymes are analogous, there are key differences in the substrate binding site. Differential placement of positively charged residues that bind the negatively charged sulfonate group of the substrate in the substrate binding region supports the proper orientation of the hydroxyl and the hydrogen group of the enzyme for proton and hydride abstraction. The sulfonate group is a crucial determinant for orienting the reactive portion of the substrates for chiral discrimination.

CHAPTER 3

STRUCTURAL ANALYSIS OF THE MET187ALA AND MET192ALA MUTANTS OF *R*-HYDROXYPROPYL COENZYME M DEHYDROGENASE

Chapter Abstract

The crystal structure of the Met187Ala and the Met192Ala mutants of the *R*-HPCDH enzyme have been determined to 1.5 Å and 2.0 Å, respectively. Both these structures were determined in the presence of the substrates *R*-HPC and NAD^+. The overall structures of these mutants are similar to the wild type *R*-HPCDH. Based on the product bound-structure of the wild type *R*-HPCDH, it was proposed that these methionine residues were involved in shielding the electrostatic interactions occurring at the active site between the substrate and the enzyme from the immediate surrounding solvent environment thereby enhancing them. Initial electron density maps of the Met187Ala structure reveal that the sulfonate moiety is not visible at the expected location thereby suggesting that product binding is altered in this mutant form of the enzyme. Superimposition of the Met187Ala mutant structure on the wild type structure reveals a huge solvent accessible pocket near the sulfonate binding site. This suggests that, non productive interaction of the arginine residues involved in sulfonate binding, with surrounding solvent molecules is possible because of the opening of a rather compact substrate binding site. The Met192Ala mutant structure clearly reveals electron density for sulfonate moiety but reveal

differences near the active site. The methionine is in close proximity to the business end of the substrate which interacts with the active site shielding the electrostatic interactions in those regions from the surrounding environment. Changing this to Ala creates a solvent accessible pocket which enables non productive interactions with surrounding solvent and thereby is not favorable for proper catalysis.

Introduction

The most extensively studied pathway of aliphatic alkene oxidation is that of propylene metabolism in *Xanthobacter autotrophicus* Py2 and *Rhodococcus rhodochrous* B276. Alkene metabolism in these organisms is initiated by an alkene monoxygenase which catalyzes the conversion of propylene to epoxypropane. The novel conversion of an epoxide (epoxypropane) to a β-ketoacid (acetoacetate) occurs by a three step process requiring four enzymes, CO_2, NAD^+, NADPH, and coenzyme M. The first step of this pathway involves the ring opening reaction catalyzed by Epoxyalkane CoM transferase which carries out the conjugation of CoM to the epoxypropane. Integral to this four enzyme pathway are two stereo specific 2-hydroxypropylCoM dehydrogenases that individually allow for metabolism of only a single epoxy propane enantiomer [96]. The *R*- and *S*-HPCDH enzymes are homologous enzymes that share a relatively high degree of sequence similarity and belong to the class of short chain dehydrogenases reductase super family. Recent work involving

amino acid modification in *R*-HPCDH enzyme has implicated the involvement of two arginine residues (Arg-196 and Arg152) in substrate binding. These residues have been proposed to support substrate binding through electrostatic interactions between the positively charged arginine side chains and the negatively charged oxygen atoms of the substrate sulfonate group. The negatively charged sulfonate group offers a convenient "handle" on the substrate molecule for binding to the active site. Presumably, similar interactions between the sulfonate binding group and positively charged amino acid side chains are involved in the binding of substrate in both *R*- and *S*-HPCDH. Based on the aforementioned site directed mutagenesis studies of *R*-HPCDH and the crystal structure of 2-KPCC, a model has been proposed for the basis of stereo selectivity in which distinct sulfonate binding sites in the *R*- and *S*-HPCDH promotes the correct orientation for hydride abstraction of their respective substrates.

In order to provide a structural basis for stereo selectivity, the crystal structure of *R*-HPCDH has been solved with product bound at the substrate binding site. This structure, along with the homology model of *S*-HPCDH has been successful in providing a structural basis for stereoslectivity. Effective hydride abstraction by short chain dehydrogenases is dependent on the correct orientation of the hydride with respect to the nicotinamide and the correct orientation of the substrate hydroxyl group with respect to the Tyr of the catalytic triad. In *R*-HPCDH the sulfonate of the substrate is bound to two

arginine side chains, Arg152 and Arg196 via electrostatic interactions. This brings about proper binding and positioning of the hydroxyl group of the substrate with respect to the active site tyrosine and the NAD^+. Comparison of this structure with the homology model of *S*-HPCDH with modeled substrate reveals that, in *S*-HPCDH, two positively charged residues (Arg205 and Lys209) could be involved in sulfonate binding. It is clear from the comparison that differential placement of positively charged residues within the substrate binding region of the enzymes provides the "switch" required for chiral discrimination.

The substrate binding site of both *R*-HPCDH and *S*-HPCDH have several other notable features that allow specific interaction of the substrate chemical groups to bring about proper binding and orientation of the substrate for catalysis. The substrate binding region appears to be a compact pocket from which solvent molecules are excluded. One end of this pocket comprises of the active site residues serine and tyrosine which are part of the catalytic triad. The other end of the pocket is lined with two arginine residues and a tryptophan residue. While the arginine residues are involved in electrostatic interactions with the sulfonate group, the tryptophan is probably acting as a back stop to prevent translation of the substrate during catalysis. An important feature of the substrate binding site is the presence of two methionine residues that appear to be flanking the sides of the substrate binding pocket. Such a compact substrate binding site is also featured in 2-KPCC, the enzyme that carries out the reductive cleavage and carboxylation of 2-ketopropyl CoM to acetoacetate and CoM in the

last step of the pathway of propylene metabolism in *X. autotrophicus*. The substrate binding pocket in 2-KPCC is strikingly similar to that of *R*-HPCDH, where one side of the pocket comprises of the active site cystienes while the other end of the pocket is lined by two arginine residues which are involved in interactions with CoM, an aromatic residue, (phenylalanine) which could act as a backstop and two methionines (Met140 and Met361) flanking the sides of the pocket. While the roles of the active site residues, the arginines and the aromatic amino acids are pretty evident form the structure, the role of methionine needs further investigation.

An analysis of the substrate binding site in context to the chemical characteristics of the substrate and its interactions with the enzyme could provide clues to the role of methionines in the dehydrogenases. The substrate molecules *R*- and *S*-HPC can be envisioned as having two chemical groups which could assist in interaction with the enzyme. First, the negatively charged sulfonate group, and second, the chiral carbon which is attached to the hydroxyl and hydrogen groups where the chemistry occurs. The sulfonate is negatively charged and the hydroxyl group attached to the chiral carbon develops a negative charge during catalysis because of the abstraction of the proton form the hydroxyl group by the deprotonated active site tyrosine residue. For effective catalysis to occur, it is important to shield these charges from each other. From the product bound-crystal structure of *R*-HPCDH, it is clearly evident that the crucial determinant of the substrate binding to the enzyme is the sulfonate group.

Major part of the enzyme-substrate interactions is contributed by electrostatic interactions between the negatively charged sulfonate group and the positively charged arginine residues. Thus, it is important to shield these interactions from the ambient solvent molecules which could compete to bind to the arginine residues via hydrogen bonding. Further, the interaction of the active site residues (Ser 142 and Tyr155) is also electrostatic in nature. There is a need to prevent non- productive hydrogen bonding interactions between the ambient solvent molecules with the active site residues so that these residues could interact with the substrate.

In the product-bound structure of *R*-HPCDH it can be seen that Met187 is in close proximity to the sulfonate group and appears to be shielding the electrostatic interactions between the sulfonate and the arginine residues from both solvent molecules and the active site electrostatic interactions. On the other hand, Met192 is in close proximity to the chiral carbon of the substrate molecule where chemistry occurs and appears to be shielding the interactions of the substrate with the active site residues form the surrounding part of the enzyme. (Figure 30)

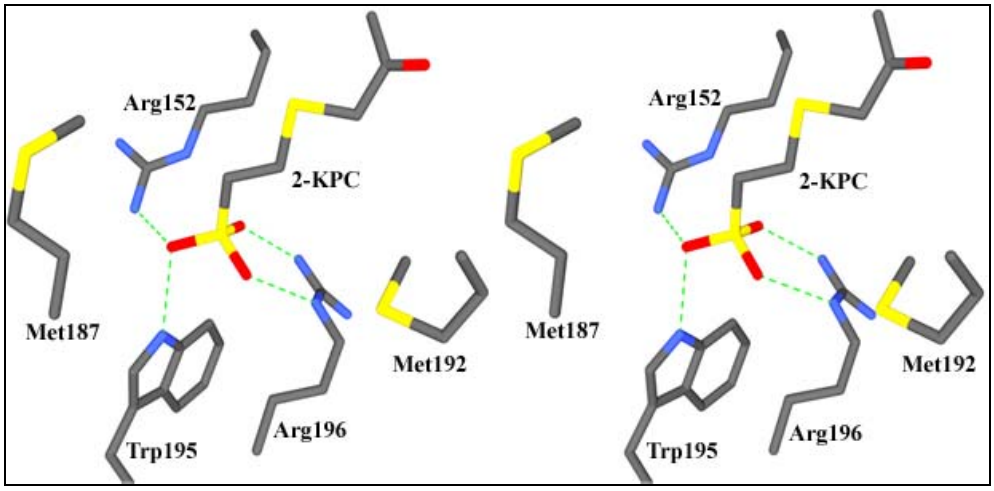

Figure 30. Key amino acid residues involved in substrate binding in the R-HPCDH enzyme.

In order to test the above hypotheses where methionines are thought to play a role in substrate binding by shielding the electrostatic interactions from the surroundings, these residues were mutated to alanine. These studies could provide some clues in to the roles played by them both in binding of the substrate and catalysis. The mutagenic and the subsequent kinetic studies have been conducted by the Scott Ensign group at Utah state University at Logan, Utah. Three different mutants; Met187Ala, Met192Ala and a double mutant where both the methionines have been converted to alanine were created to test their effects on catalysis and substrate binding.

Kinetic analysis of these mutants has revealed interesting results. Both K_{cat} and K_m are affected when the methionines are substituted by alanines in the *R*-HPCDH enzyme. When the enzyme assay was conducted with the substrate *R*-HPC, it was seen that for the M187A mutant The K_m was ~88 fold higher and K_{cat} was 1.8 times lower than the wild type. In the case of the M192A mutant K_m

was not affected as much as it was with the M187A (21.46) but the kcat was drastically lower (~22 fold). The double mutant seems to have lost the catalytic ability with negligible K_{cat} values and K_m values were very similar to the M187A mutant. (Table 4) From the above studies it can be seen that the methionines are crucial for both substrate binding and catalysis.

Table 4. Kinetic analysis of the Met187Ala and Met192Ala mutants of the R-hyrdoxypropyl CoM dehydrogenase.

Enzyme	Km [mM]	Change in Km (x-fold)	Vmax [U/mg]	K_{cat} [sec^{-1}]	*Kcat* (x-fold lower)
R-HPCDH WT	0.044	1.00	52.83	26.17	1.00
M187A	3.682	82.76	29.26	14.49	1.8
M192A	0.955	21.46	2.42	1.19	21.8
Double	3.885	87.32	0.01	0.0042	6220.4

Structural analyses of these mutants could provide more insights into the roles of these methionines. With the structure of *R*-HPCDH already in hand, obtaining the structure of these mutants would be trivial if they crystallize in the same space group. The present study involves the structural characterization of some of these mutants to study the role of methionines in catalysis and substrate binding.

Materials and Methods

The crystals of the Met187Ala and Met192Ala were obtained under the same conditions as the wild type. Prior to crystallization the substrates NAD^+ and *R*-HPC were added to the protein at a concentration of 15 and 20 mM,

respectively. The crystals grew in 2 weeks and were 0.2 x 0.3 x 0.2 mm in size (Figure 31). These crystals were harvested on rayon loops and were flash frozen in liquid nitrogen and data sets were collected at SSRL beam line 9-2. The data sets were processed using HKL2000 [81] and the data statistics for each of the mutants have been reported in Tables 5 and 6, respectively. The crystals belonged to the space group $P2_1$ and had identical cell parameters as the previously characterized crystals of the wild type *R*-HPCDH. Since the cell parameters were identical, the structure solutions for these mutants were directly obtained by using the *R*-HPCDH wild type structure as the starting model. The refinement of the structures was carried out using the CNS suite of programs [84]. Several cycles of refinement alternated by model building using the program O [83] resulted in final models which obey reasonable stereochemistry with 100% of residues occupying allowed regions of a Ramachandran Plot calculated using PROCHECK [86]. The current refinement statistics for both the structures are provided in Tables 7 and 8, respectively.

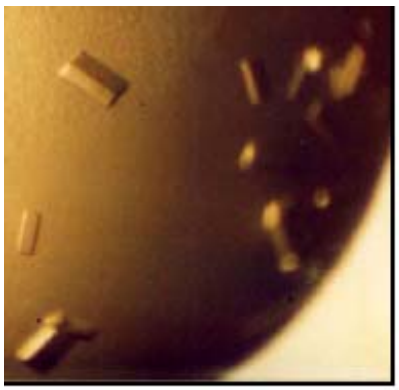

Figure 31. Crystals of the Met187Ala mutant of *R*-HPCDH.

Table 5. Data statistics for the *R*-HPCDH mutant Met187Ala Co-crystallized with *R*-HPC.

Space group	$P2_1$
Cell dimensions	a = 64.678 Å b = 110.176 Å c = 68.894 Å $\alpha = \gamma = 90^o$ $\beta = 93.855^o$
Resolution (Å)	20.00-1.46
Completeness (%)	97.9(93.7)
Observed reflections	1163058
Unique reflections	161616
I/σ	48 (7.5)
R_{merge}[a]	0.042(0.301)

Table 6. Data statistics for the *R*-HPCDH mutant Met192Ala Co-crystallized with *R*-HPC.

Space group	$P2_1$
Cell dimensions	a = 64.720 Å b = 110.812 Å c = 68.866 Å $\alpha = \gamma = 90^o$ $\beta = 93.710^o$
Resolution (Å)	20.00-2.08
Completeness (%)	98.9(96.7)
Observed reflections	503556
Unique reflections	65222
I/σ	30.1(15.5)
R_{merge}[a]	0.111(0.264)

Numbers in parentheses indicate values for the highest resolution bin.

a. $R_{merge} = \Sigma hkl \Sigma i|Ii - >|/\Sigma hkl \Sigma i|<I>|$, where Ii is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.

Table 7. Refinement statistics for the *R*-HPCDH mutant Met187Ala Co-crystallized with *R*-HPC.

Resolution (Å)	20.00-1.46
R_{free}	0.201
R_{cryst}	0.181

Table 7. Refinement statistics for the *R*-HPCDH mutant Met187Ala Co-crystallized with *R*-HPC-continued.

No. Non hydrogen atoms	
Protein	4004
Cofactor	176
solvent	419
RMSD from target values	
Bond lengths (Å)	0.012
Bond angles (°)	1.55
Average isotropic B factor	
All atoms	18.71
Protein main chain	13.71
Protein side chain	18.41
NAD^+	47.25
Water	19.05

Table 8. Refinement statistics for the *R*-HPCDH mutant Met192Ala Co-crystallized with *R*-HPC.

Resolution (Å)	20.00-2.08
R_{free}	0.231
R_{cryst}	0.201
No. Non hydrogen atoms	
Protein	4004
Cofactor	176
solvent	243
RMSD from target values	
Bond lengths (Å)	0.014
Bond angles (°)	1.59
Average isotropic B factor	
All atoms	15.39
Protein main chain	19.66
Protein side chain	17.33
NAD^+	50.25
Water	15.05

Results and Discussion

Overall Structure of the Mutants Met187 and Met192

The tertiary structure of the mutants is similar to that of the wild type *R*-HPCDH. In the Met187Ala mutant structure data is good up to 1.46 Å and have allowed calculation of very high quality electron density maps. Initial analysis of electron density maps of the Met187Ala structure indicates that the NAD^+ molecule is well ordered (Figure 32). It can be recalled that the electron density around the second ribose and the nicotinamide moiety of the NAD^+ was not well resolved in the previously determined wild type *R*-HPCDH structure. In the wild type structure the nicotinamide and the ribose moiety were positioned based on the orientation of those groups in the well characterized enzyme mannitol dehydrogenase whose structure is very similar to that of *R*-HPCDH. Also, it was taken into consideration that the nicotinamide and the ribose moiety could be positioned in only certain orientations to bring about proper catalysis. The present structure not only reveals clear electron density for the entire ribose molecule and a part of the nicotinamide but also confirms that the positioning of those groups in the previous structure was correct and well thought out.

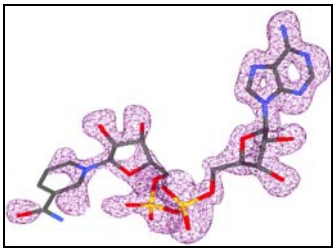

Figure 32. Electron density contoured at 1σ cutoff around NAD^+ in the structure of R-HPCDH Met187 Ala mutant.

Recently, site directed mutagenesis studies involving the Met187 and Met192 residues and kinetic studies on the mutant enzymes have revealed the importance of these residues for catalysis. (unpublished results, Scott Ensign). When Met187 and Met192 were mutated to alanine, both K_m and K_{cat} were altered for the substrate *R*-HPC. While Met187Ala brought about a significance increase in km, Met192Ala mutant enzyme had very low K_{cat}. Crystal structures of these two proteins are determined and they provide insights to hypothesize their role in substrate binding and catalysis.

The Role of Met187 in Substrate Binding and Catalysis

Crystal structure of Met187Ala reveals clear density corresponding to an alanine rather than a methionine at residue number 187 confirming the mutation. Initial electron density maps calculated at the current stage of refinement and contoured at 1σ cutoff reveal that there is no density corresponding to the sulfonate group at the sulfonate binding site agreeing with the kinetic results that this mutation has large effects of binding of the substrate (K_m). It can be seen from the maps that this position has now been occupied by density corresponding to solvent molecules (Figure33).

Superimposition of the wild type structure on the mutant structure was carried out and a substrate molecule (*R*-HPC) was modeled at the active site and the superimposition reveals the changes brought abut by the mutation. This could provide insights in to the structural changes brought about by the mutation in the substrate binding pocket. It should be recalled here that Met187 serves to

shield the sulfonate group of the substrate from electrostatic interactions with the solvent and also protects the sulfonate group from active site chemistry.

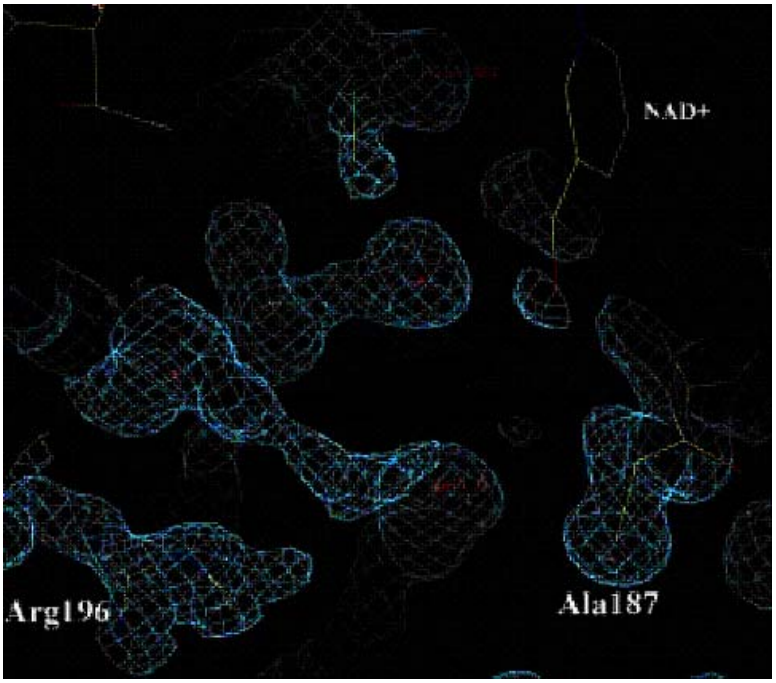

Figure 33. Electron density contoured at 1σ cutoff at the sulfonate biding region in the Met187Ala mutant structure of *R*-HPCDH.

In the wild type *R*-HPCDH, Met187 is positioned very closely to the sulfonate group and defines the sulfonate binding pocket. When the Met187 mutant structure is superimposed on the wild type *R*-HPCDH structure, it can be seen that due to the mutation of Met187 to a very small amino acid, alanine, there is a large solvent accessible pocket created around the sulfonate binding region which was previously well shielded from the surroundings (Figure 34). This could allow non-specific interactions of the arginine residues with surrounding solvent molecules or molecules such as glycerol or Tris introduced during crystallization of the enzyme and thereby hinder constructive interactions

of the sulfonate moiety with the positively charged residues. Another important feature which came into light form the superimposition of the wild type structure on the Met187Ala structure is that, the side chain of Phe149 has a different conformation in the mutant structure which suggests a possible role in substrate binding. From the wild type structure it is seen that the Phe149 is in close proximity to Met187 forming a part of the sulfonate binding pocket shielding the substrate from behind the active site. Changing Met187 to Ala allows the side chain of Phe149to be in a different conformation creating a solvent accessible pocket.

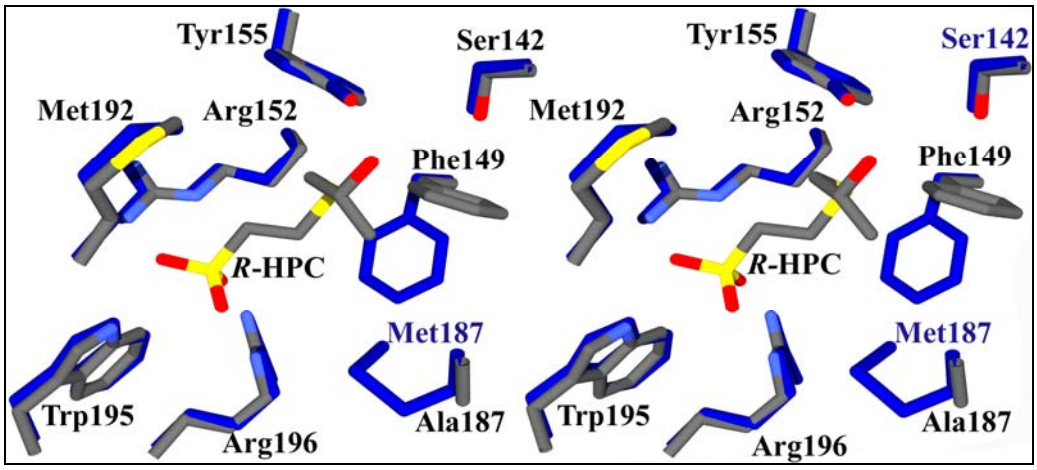

Figure 34. Structural depiction of the superimposition of the substrate binding pocket of wild type *R*-HPCDH on the Met187Ala binding pocket.

The Role of Met192 in Substrate Binding and Catalysis

Kinetic studies on Met192 reveal that the mutation of this residue to alanine does not cause a drastic change in the K_m but significantly affects *K*cat. The crystal structure of Met192Ala has been determined recently to 2.08Å and the initial maps reveal the presence of alanine in place of Met192 and thus

confirm the mutation. These maps also reveal a clear electron density for the sulfonate moiety of the substrate (Figure 35). This supports the kinetic experiments and suggests that Met192 does not have major role in sulfonate binding. The density around the hydroxyl and the methyl group of the substrate is not clear.

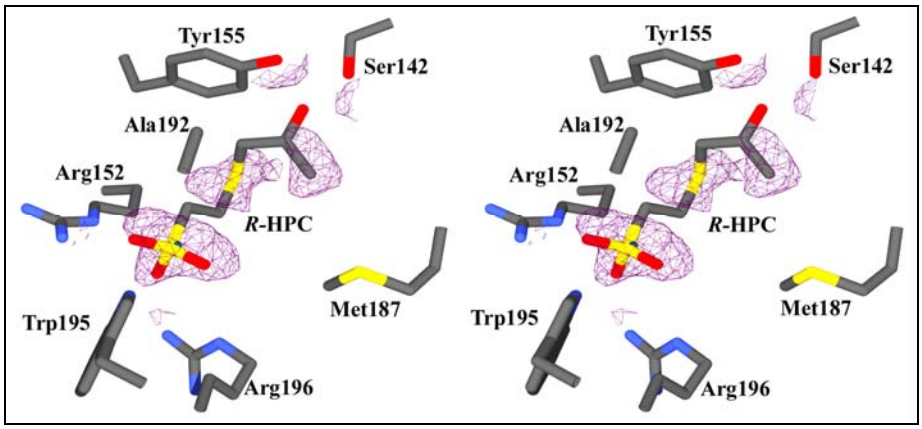

Figure 35. Electron density contoured at 1σ cutoff at the sulfonate biding region in the Met187Ala mutant structure of *R*-HPCDH.

A hypothesis for the role of Met192 can be drawn by comparing the structure of the mutant enzyme and the wild type *R*-HPCDH by superimposition of the two structures (Figure 36). Met192 in the wild type enzyme is positioned such that it shields the part of the substrate molecule involved the dehydrogenation-chemistry i.e. the hydroxyl and the hydrogen groups. Prior to substrate binding this residue appears to be responsible for defining the substrate binding pocket near the catalytic triad and shield the negative charge on the active site tyrosine from ambient solvent molecules thereby creating a solvent inaccessible pocket. This ensures that specific interactions of the active site residues can occur with

the substrate hydroxyl group thereby bringing about effective catalysis. Superimposition of the structure of the wild type enzyme on the Met192 mutant reveals that, in the absence of Met192, a solvent accessible pocket is created around the active site. The solvent molecules can compete with the hydroxyl and hydrogen group of the substrate and hinder the reaction. This structure supports the kinetic data that Met192 is not involved in substrate binding but plays an important role in catalysis.

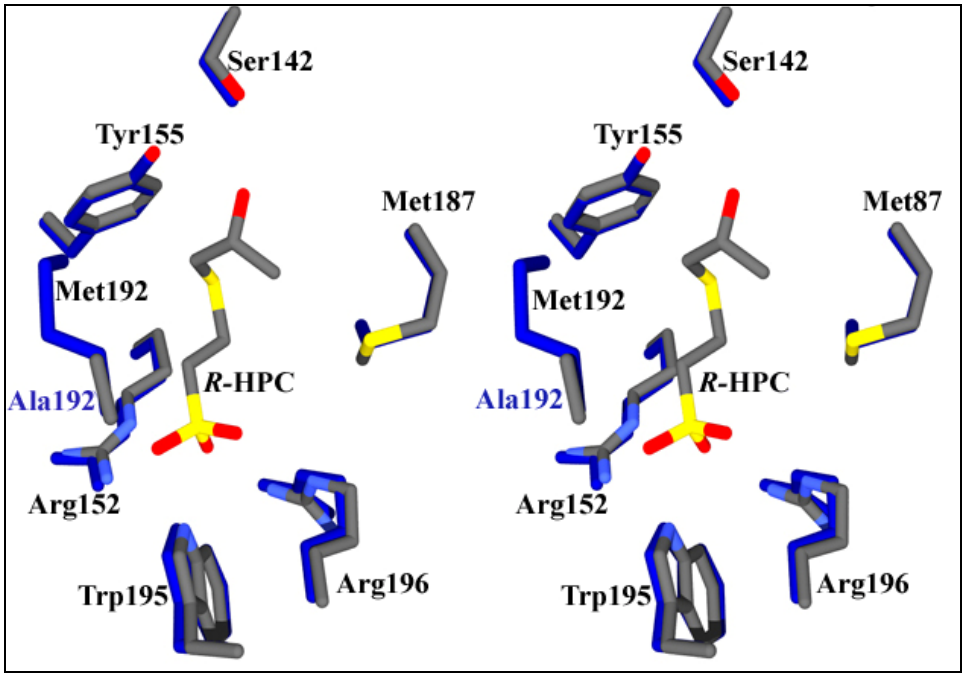

Figure 36. Structural depiction of the superimposition of the substrate binding pocket of wild type *R*-HPCDH on the Met192Ala binding pocket.

Summary and Conclusions

The crystallographic structures of Met187Ala and Met192Ala and their comparison with the wild type structure reveals the roles played by these residues in

substrate binding and catalysis respectively. These residues are positioned in close proximity to the substrate binding pocket. While methionine 187 defines the sulfonate binding site and shields the electrostatic interactions of the arginines and the sulfonate group and enhance them. Also, its interactions with the side chain of Phe149 provides for a solvent protected sulfonate binding pocket. Methionine 192 on the other hand defines the substrate binding pocket around the catalytic triad and is placed in close proximity to the hydroxyl and hydrogen group of the substrate. Since the substrate *R*-HPC has a negatively charged sulfonate group at one end and a polar group at the other it is very important to protect those groups from interacting unproductively with the surrounding solvent molecules. By strategically positioning these methionines around the substrate binding pocket it has been ensured that the interactions between the substrate and the enzyme are enhanced to bring about proper catalysis.

CHAPTER 4

ROLE OF COENZYME M IN ALKENE METABOLISM

Chapter Abstract

Coenzyme M was discovered during studies aimed at providing insights into the process of methanogenesis. Uniquely simple, coenzyme M (2-mercaptoethanesulfonate) so named for its role as methyl carrier in all methanogenic archaea is the smallest known organic cofactor. It was thought that this cofactor was exclusive to methanogenesis until the late nineties when it was discovered that coenzyme M was a key cofactor in the pathway of propylene metabolism in the gram negative soil microorganism *Xanthobacter autotrophicus* Py2. A four step pathway converts propylene to acetoacetate that can be funneled into the central metabolism allowing *X. autotrophicus* to utilize propylene as sole carbon source. Short chain hydrocarbons lack the chemical groups necessary for the specific binding to enzymes in specific orientations. The conjugation of substrates and intermediates to coenzyme M provides a chemical handle for orienting compounds for site specific redox chemistry and stereo specific catalysis. We have determined the three-dimensional structures of several of the enzymes in the pathway of propylene metabolism in defined states providing significant insights into both the enzyme mechanisms and the role of coenzyme M in this pathway. These studies provide the structural basis for understanding the efficacy of coenzyme M as a handle to direct organic

substrates that have limited innate chemical characteristics at the active site of enzymes that catalyze highly selective catalysis.

Introduction

Coenzyme M [26], the smallest known organic cofactor (ethane mercaptosulfonic acid – Figure 37), was isolated and characterized by McBride and Wolfe in 1971 [36, 97] and demonstrated to have a role as a methyl carrier in the process of methanogenesis. The metabolic pathways for methane production, regardless of growth substrate, converge via the activity of a variety of methyltransferases on a common metabolic reaction in which methyl substituted CoM (methyl-CoM) is reductively cleaved by the activity of Ni- dependent methyl CoM reductase to form methane. Although the pathway for the biosynthesis is not been completely elucidated, this is a very active research and hypothetical pathways have been proposed [98, 99]. In addition more recently, several of the enzymatically catalyzed steps being experimentally characterized. CoM was thought to be exclusively associated with methanogens and for many years was used as a biomarker to probe for the presence of Methanogenic Archaea in environmental samples.

In the late nineties, to some surprise, it was discovered that CoM was a cofactor in the pathway of propylene metabolism in *X. autotrophicus [31]*. *Xanthobacter* is one of several organisms than can use alkenes and epoxides as sole sources of carbon for growth. Since these epoxides are often toxic and lead

to deleterious effects by forming conjugates with various biomolecules, microorganisms that encounter high concentrations of these compounds must convert them to compounds less detrimental. This can be accomplished by either metabolism or detoxification. Detoxification typically involves ring opening by the nucleophilic attack of either glutathione or water to form covalent adducts by the activity of glutathione S-transferase or epoxide hydrolases and discarded or further metabolized. In contrast, the alkene-oxidizing bacteria are unique in being able to carry out the productive catabolism of aliphatic epoxides by their conversion to central metabolites that meet the carbon and energy needs of the bacteria. In this process, alkene epoxidation is followed by conjugation with CoM and subsequent carboxylation to the metabolite acetoacetate in a pathway in which the chemical characteristics of CoM play a key role in the binding and orientation of conjugated substrates for catalysis.

It is intriguing that CoM is found to be the cofactor for epoxide carboxylation, not only because CoM was thought to be an exclusively methanogenic coenzyme but also given the availability of other thiols like glutathione, lipoic acid, cysteine, or homocysteine that could participate in this type of chemistry. CoM is one of the very few sulfonic acids found in nature and is the smallest organic cofactor known with a molecular weight of 142 daltons. Also termed 2-mercaptoethanesulfonic acid, it is composed of an ethyl spacer between thiol and sulfonic acid functional groups (Figure 37). The advantage of using CoM in the epoxide carboxylation pathway is two-fold.

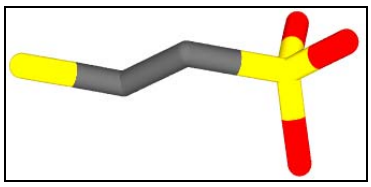

Figure 37. Structure of coenzyme M (mercaptoethane sulfonic acid).

Firstly, the small size of CoM allows the formation of 2-hydroxypropylCoM conjugate whose size is highly manageable as opposed to bulky glutathione-alcohol conjugates formed by glutathione S-transferases, for stereo selective dehydrogenation and reductive carboxylation in the subsequent steps. Secondly, the negatively charged sulfonate moiety of CoM provides a group on the substrate molecule to which it has been conjugated that can be utilized as a handle for orienting them at the active sites of the enzymes for site specific redox chemistry and stereo specific catalysis. This is a very important feature in the context of alkene metabolism because short chain alkene and epoxide substrates lack and special chemical groups that could confer specificity.

The unique features of CoM have been exploited in methanogenesis where it is used as a methyl group carrier. The involvement of CoM in epoxide carboxylation is similar to its role in methanogenesis. There are some interesting similarities between the pathways of epoxide carboxylation (Figure 38 A) and methanogenesis (Figure 38 B) with respect to CoM utilization. In both cases, the initial group transfer involves *S*-alkylation of CoM to form a thioether intermediate. In case of methanogenesis, this reaction is carried out by a variety of methyl transferases. In order to conjugate the CoM with the methyl group, the

cofactor must first be activated as a nucleophile by deprotonation of the sulfhydryl to generate the thiolate anion.

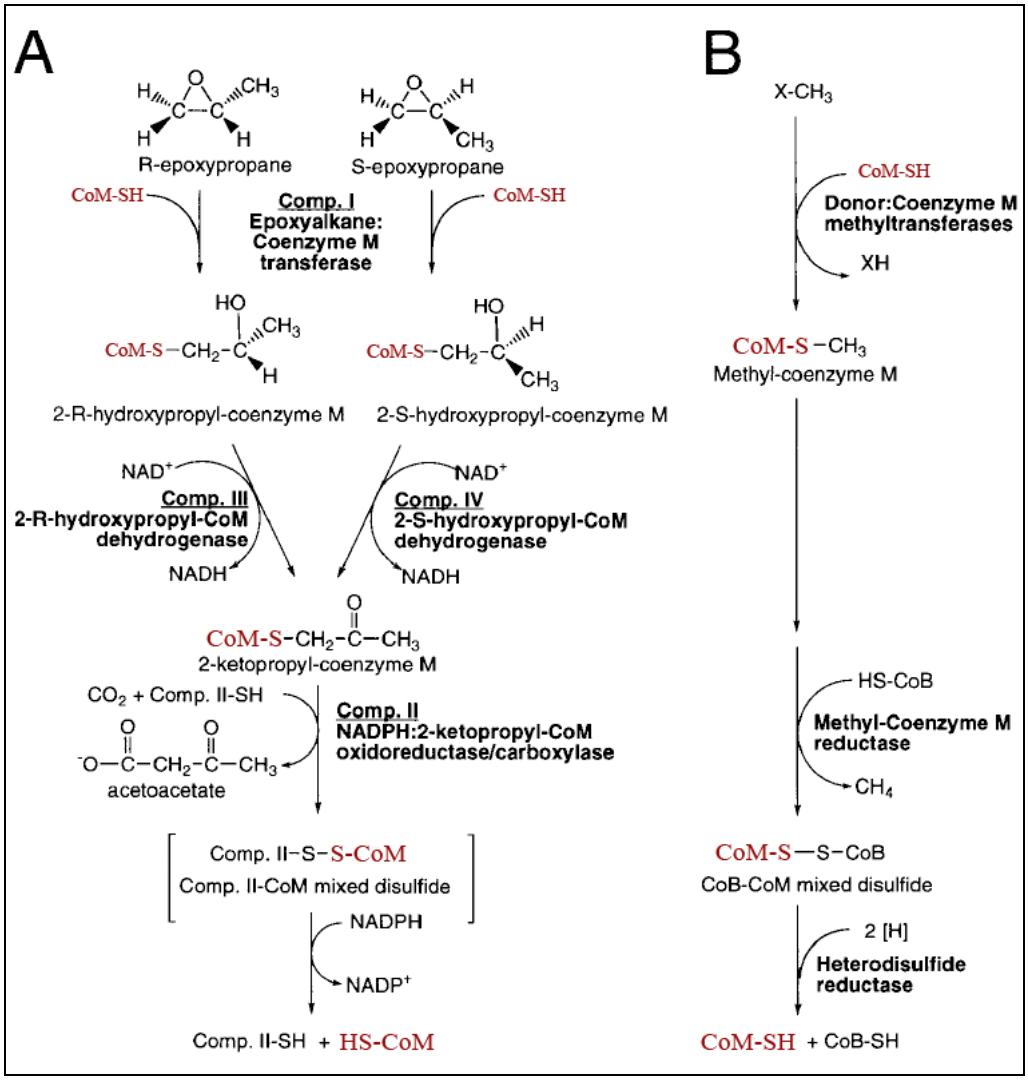

Figure 38. Comparison of CoM usage in epoxide carboxylation (A) and methanogenesis (B).

It has been shown in methanogenic methyl transferases that zinc plays and integral role in the activation of the thiol group of CoM for acceptance of methyl group from various donors [100]. Analogously, the zinc in EaCoMT in the

epoxide carboxylation pathway functions in the activation of CoM as well, in this instance for attack on the electrophilic epoxide substrate [48]. In both methanogenesis and epoxide carboxylation, a reductive dealkylation of a CoM thioether occurs, in the former system to generate methane, in the latter to form an enolate that undergoes carboxylation. Apart from *Xanthobacter autotrophicus* and *Rhodococcus rhodochrous*, other organisms where CoM dependent alkene metabolism has been identified, isolated and studied in detail includes strains of *Mycobacterium rhodesiae*[101], *Nocardiodes* [102], *Pseudomonas putida* and *Ochrobactrum* [103]. Several strains of these microorganisms are able to grow on ethene and vinyl chloride. The first step in the ethene and vinyl chloride degradation in these organisms is known to be a monoxygenase reaction producing epoxyethane and chlorooxirane, respectively [101, 104]. In the next step EaCoMT has been shown to be involved in the ring opening reaction. The downstream reactions are not well understood in these organisms, but a hypothetical pathway has been proposed.

The discovery of CoM as a cofactor of a very specialized hydrocarbon oxidation pathway and the previously known fact that it is involved in methanogenesis suggests that it may be expressed and used by bacteria/archea which carry out special type of metabolism such as epoxide carboxylation and methanogenesis, respectively. Rather than using a broad specificity detoxification enzyme, *X. autotrophicus* uses epoxide carboxylation which utilizes a unique cofactor to form a growth supporting central metabolite.

Improving our understanding the role of CoM as a carrier molecule in alkene oxidation and methanogenesis is greatly facilitated through investigating of three dimensional structures of the enzymes of the aforementioned pathways complexed with substrates, products, and/or inhibitors. Structures as a complement to detailed kinetic studies can be extremely informative for establishing models for enzymatic mechanism that can be approached experimentally. The three dimensional structures of some of the enzymes of the propylene oxidation in *X. autotrophicus* have been solved [66, 96]. These structures have provided the structural basis for understanding the efficacy of CoM as a cofactor by revealing the interactions between CoM and the amino acid side chains of the enzymes defining a key role for CoM in specifically binding and orientating substrates for substrate carboxylation and stereo selective catalysis. In all these enzymes, strategically placed positively charged amino acids (arginine and lysine) interact with the negatively charged sulfonate group of the substrate, thereby orienting it properly for catalysis. A common pattern of amino acid-substrate sulfonate interaction involved in all these enzymes can be seen which suggests a structural signature for CoM interactions. In the present study structures (where available) and homology models of the enzymes which bind CoM or its conjugates are analyzed in detail in order to find such a structural signature. These analyses could provide a structural basis for the usage of CoM as a unique cofactor of epoxide carboxylation.

Epoxyalkane CoM Transferase (EaCoMT): Directing the CoM Thiol Group for Zinc Mediated Activation

The EaCoMT catalyzed reaction is a very important step in alkene and chlorinated alkene metabolism because it couples the epoxide ring opening to the conjugation of the resulting alcohols with coenzyme M. EaCoMTs belong to the class of zinc dependent thiol activated alkyl transferases [39]. Other examples of this family are ADA protein [43, 44], betaine:homocysteine methyltransferase [45], and cobalamin-independent methionine synthases (MetE) [100, 105], and various methanogenic methyl transferases (MtaA, MtbA and MtsA).

EaCoMTs were first identified in *X. autotrophicus* [32, 33] and *R. rhodochrous* [34] as the enzyme which catalyzes the addition of CoM to *R*-or *S*-epoxypropane. Recently, EaCoMTs have been identified in other organisms such as *Mycobacterium*, *Nocardiodes, Pseudomonas and Ochrobactrum* which metabolize ethene and vinyl chloride. EaCoMTs from these organisms share high degree of sequence similarity and identity between them. Multiple sequence alignment (Figure 39 A) and phylogenetic tree (Figure 39 B) based on the EaCoMT primary sequences indicate that the *Pseudomonas* and *Ochrobactrum* genes cluster with the *Mycobacterium.* Although *Nocardiodes* and *Rhodococcus* cluster together they are very closely related to the *Mycobacterium. X. autotrophicus* is the quite different from all the above strains and it clusters separately.

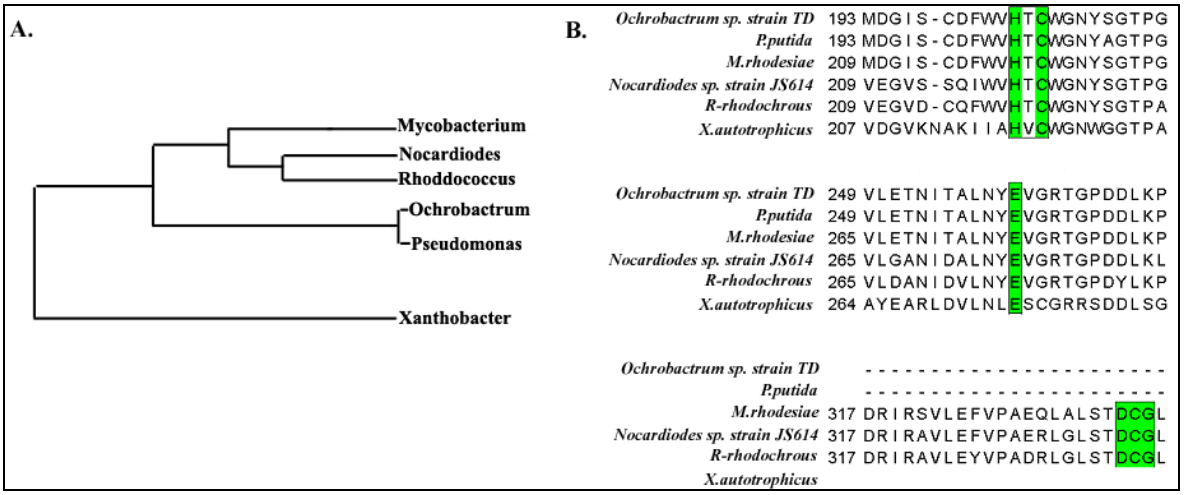

Figure 39. (A) Multiple sequence alignment carried out using ClustalW [106] and (B) phylogenetic tree constructed using Phylip [107] indicating the relative sequence similarity of epoxyalkane coenzyme M transferases from microorganisms in which they have been implicated to occur by genome analysis.

Multiple sequence alignment of these EaCoMTs reveals the HXCXnC motif which is common in transferases which use zinc to activate a thiol group. In all these enzymes, zinc is coordinated by two cystienes and a histidine acting as permanent ligands. Whereas the *Xanthobacter, Rhodococcus, Nocardiodes* and *Mycobacterium* have the full component of the motif, the *Pseudomonas* and the *Ochrobactrum* enzymes have only the histidine and the first cysteine. Apart from the zinc binding motif there are other regions all along the sequences which are highly conserved amongst these enzymes.

Amongst all EaCoMTs the one from *X. autotrophicus* is studied most extensively [48, 108]. It catalyzes the addition of CoM to either *R* or *S* epoxypropane forming 2-*R* and 2-*S* hydroxypropyl CoM, respectively (*R*-HPC and *S*-HPC for short). It is a hexameric protein containing 1 Zn per subunit.

EaCoMT is highly specific for using CoM as the organic thiol substrate. When a range of other thiols were tested, only 3-mercaptopropionate, 2-mercaptoethanol and cysteine served as substrates with very low affinities and specific activities suggesting specific interactions between CoM and the enzyme [48].

Multiple sequence alignment with other zinc dependent alkyl transferases has shown that EaCoMT from *X. autotrophicus* is very similar to that of zinc-dependent methionine synthases (MetE). Extensive biochemical, spectroscopic and crystallographic studies have been carried out on various methionine synthases from organisms such as *Thermotoga maritima* and *Arabidopsis thaliana*. When compared to the sequences of methionine synthases whose three dimensional structure has been determined, the primary amino acid sequence EaCoMT from *X. autotrophicus* is most similar to that from *T. maritima.* Sequence alignment of EaCoMT with *T. maritima* MetE aligns Cys220, Cys341 and His218 perfectly with the well characterized zinc binding motif of MetE involving Cys643, Cys726 and His641 [28, 109] (Figure 40). Mutagenesis experiments have confirmed Cys220 as one of the ligands [48] and extended X-ray absorption fluorescence spectroscopy (EXAFS) analysis support the 2S+ 2N/O ligand environment [108]. While histidine forms the1N/O environment, it the second ligand is assumed to be a water molecule. EXAFS have revealed that, upon CoM binding to the enzyme, the ligand environment changes to 3S + 1N/O with two cystienes and the CoM forming the 3S environment and the histidine forming the 1N/O environment. Similar EXAFS studies have been conducted on

the *T. maritima* MetE revealing that upon homocysteine binding, the zinc environment changes from 2S + 2 N/O to 3S + 1N/O. In this case, two cystienes and a histidine make up for two sulfurs and one nitrogen environment. The oxygen ligand comes from a glutamate carboxylate.

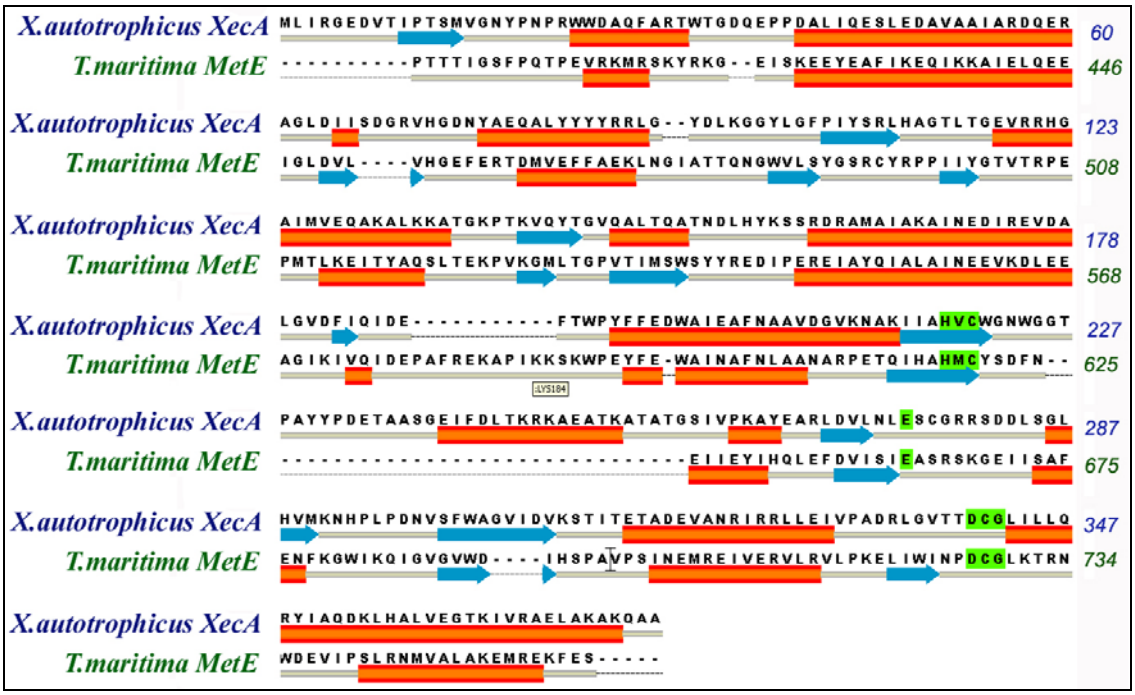

Figure 40. Alignment (using discovery Studio 1.6) of MetE (zinc-dependent methionine synthase) from *T. maritima* and XecA (epoxyalkane coenzyme M transferase - EaCoMT) from *X. autotrophicus* Py2 with zinc binding motif highlighted in green.

Crystal structure of MetE from *T. maritima* confirms the EXAFS studies and reveals that zinc is coordinated by Cys620 and Cys704 and His618 as permanent ligands whereas a glutamate carboxylate forms the fourth ligand [110]. The substrate bound structure is also available and reveals that upon homocysteine binding, the thiol group of homocysteine coordinates the zinc and the glutamate carboxylate moves away from the zinc thereby increasing the

coordination distance. The crystal structure of EaCoMT has not yet been determined, nevertheless, high sequence similarity between the *X. autotrophicus* EaCoMT and MetE has allowed us to construct a homology model (Figure 41 A) of the former using the crystal structure of the latter. The homology model has allowed us to visualize a putative zinc binding site in *X. autotrophicus* based on the crystal structure of MetE from *T. maritima* (Figures 41 B & 41 C, respectively).

Although it has been assumed form EXAFS studies that a water molecule could be the exchangeable ligand, homology model shows that glutamate is more likely the fourth ligand. EXAFS provide information only about the coordination environment and not about the particular amino acid residue involved in coordinating the zinc. It should be noted that this glutamate residue is conserved in EaCoMTs in all the aforementioned eubacteria. The fact that glutamate could be possibly the fourth ligand might have been overlooked because there are no mutagenesis studies conducted on this conserved residue. The first clue that glutamate is possibly the fourth ligand comes from these homology models and calls for further investigation. Mutagenesis experiments performed on this glutamate residue would confirm its role in zinc binding. The zinc binding sites of *Pseudomonas* and *Ochrobactrum* are different with only one cysteine, histidine and a glutamate. The substrate binding site is conserved in all the EaCoMTs. One end of the substrate binding pocket has the zinc bound to the

HXCXnC motif. The other end comprises of Trp, Phe and Tyr forming a hydrophobic pocket.

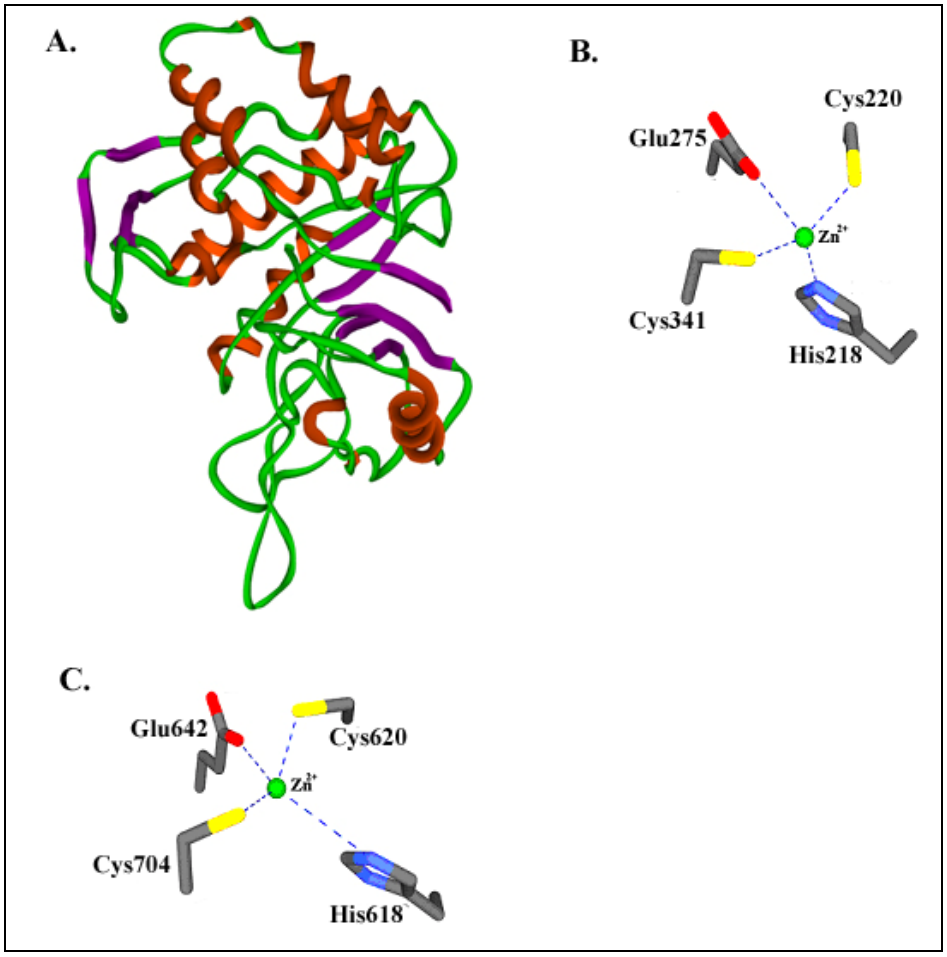

Figure 41. (A) Homology model of the structure of XecA/EaCoMT from *X. autotrophicus* Py2 based on the structure of MetE generated using CPH models [88] server. Structural representations comparing the zinc binding sites of (B) EaCoMT and (C) MetE. All figures were created using SWISS PDB VIEWER [90]and rendered by POVRAY[91].

Also found in the substrate binding site are two positively charged residues, arginine and lysine which are conserved in all of the organisms indicated. The details of the specific amino acid residues that could be involved in zinc and CoM binding in the various EaCoMTs are summarized in (Figure 42 A). The

arginine residue in the *Rhodococcus, Mycobacterium, Ochrobactrum, Pseudomonas* and *Nocardia* is conserved with respect to both its position in the sequence and three dimensional orientations in the tertiary structure. Interestingly the same arginine residue in *X. autotrophicus* EaCoMT is only structurally conserved. The arginine residue in the *X. autotrophicus* enzyme is far away in the protein sequence from the conserved arginines seen in the other EaCoMTs. It is intriguing how the architecture of the substrate binding sites is structurally conserved even if the residues are present on different regions of the protein sequences.

In the crystal structure of MetE it can be clearly seen that the homocysteine coordinates the zinc ion via the sulfhydryl group. The distance between the thiol group and the zinc ion is ~3.2Å. In order to envision the substrate-enzyme interaction and alkyl transfer reaction in the EaCoMT, the MetE structure along with the bound homocysteine was superimposed on to the structure of EaCoMT obtained from homology modeling. This put the zinc exactly at the putative zinc binding site consisting of His218, Cys 220, Cys341 and Glu275. Also, the superimposition resulted in positioning of the homocysteine molecule close to zinc such that it could serve as a ligand to the zinc ion. The distances between the zinc ion and its ligands namely the two cystienes, histidine and the thiol group were consistent with the ones observed in

A.

	His	Cys	Cys	Glu	Arg	Lys
X. autotrophicus	218	220	341	275	104	309
R. rhodochrous	219	221	337	276	279	305
M. rhodesiae	219	221	337	276	279	305
Nocardiodes sp. strain JS614	219	221	337	276	279	305
P. putida strain AJ	203	205	-----	260	263	289
Ochrobactrum sp. strain TD	203	205	-----	260	263	289

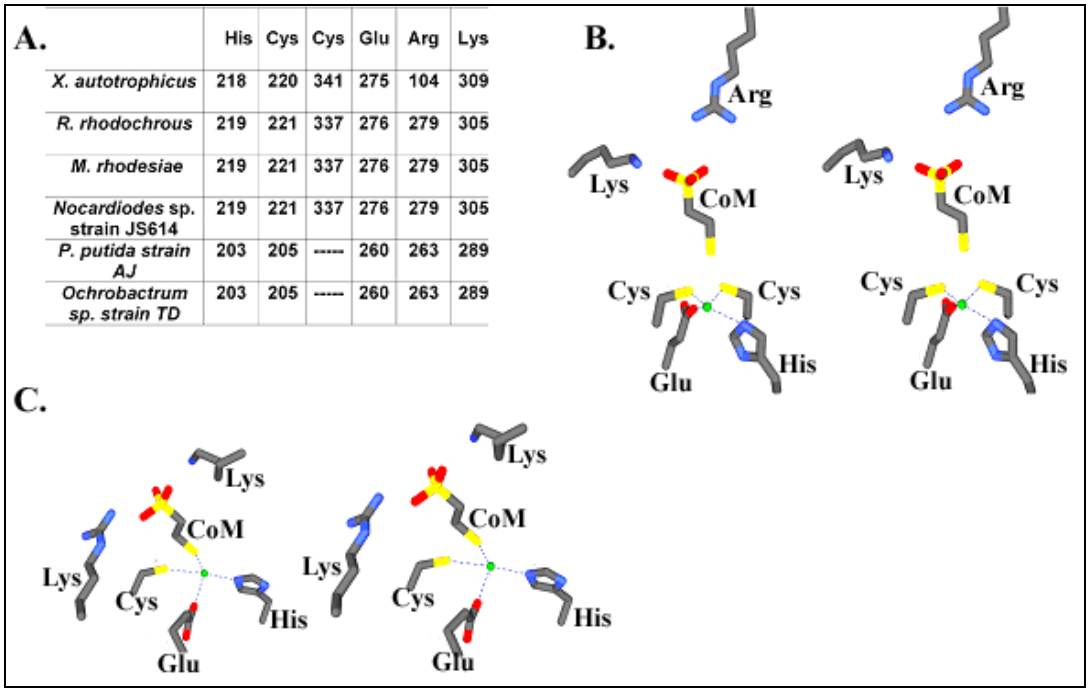

Figure 42. (A) Amino acid residues involved in CoM and zinc binding in EaCoMTs from *X. autotrophicus*, *R. rhodochrous*, *M. rhodesiae* and *Nocardiodes sp. Strain JS614*, *P. putida and Ochrobactrum* sp strain TD (B). Structural representation of the key residues of the zinc binding sites in *X. autotrophicus*, *R. rhodochrous*, *M. rhodesiae* and *Nocardiodes sp. Strain JS614* and (**C**). P.putida and *Ochrobactrum* sp strain TD. Docking of substrates were carried out using the program O[83]. In-situ minimization of the models with the bound substrates was carried out using the DS modeling suite of programs which uses a CHARMm minimization protocol.

the crystal structure (~3.2Å). Considering the position of the thiol group of homocysteine in the EaCoMT model as a starting point, a CoM molecule was modeled by manual docking using the program O, in the homocysteine binding site such that the thiol of the CoM lies exactly at the same position as the thiol of the homocysteine. Following this, rigid body movement was carried out on the CoM molecule to see if the CoM sulfonate group can be interacting with any of the structurally conserved positive residues in the substrate binding site. In situ

ligand minimization was carried out to remove any bad contacts in the models. The generalized substrate binding sites of *Xanthobacter, Rhodococcus, Nocardiodes* and *Mycobacterium* with the bound CoM are shown in Figure 42 B. The zinc and CoM binding in the *Pseudomonas* and the *Ochrobactrum* enzymes are shown in Figure 42 C. The minimized models show that the arginine and lysine residues could possibly be involved in electrostatic interactions with the sulfonate group of the CoM. These residues approach the CoM from both sides and could be involved in electrostatic interactions with the sulfonate group. The distances between the sulfonate oxygens and the side chain nitrogens of the arginine and lysine residues are similar to the previously observed distances between these groups in crystal structure of *R*-HPCDH and 2-KPCC where similar kind of interactions described in detail in their respective sections. From the crystal structures of the other enzymes in the pathway such as *R*-hydroxypropyl CoM dehydrogenase (*R*-HPCDH) and 2-ketopropylCoM carboxylase/oxidoreductase (2-KPCC), it is known that the substrates R-HPC and 2-KPC are bound by the respective enzymes through specific interactions between the negatively charged sulfonate group of the substrates and positively charged residues of the enzymes namely, arginine and lysine. These structures which are discussed in-depth in the following sections of the chapter, reveal that the positively charged residues are positioned in such a way that interaction of sulfonate group of the CoM with these residues bring about proper binding and

orientation of the alcohol and keto group of the substrates in *R*-HPCDH and 2-KPCC for dehydrogenation and thio-ether bond cleavage respectively.

In EaCoMT, the distance between the CoM sulfonate group and the positively charged residues gives insight into the specificity of EaCoMT towards CoM and not other thiols. Studies have shown that when other thiols such as 3-mercaptopropionate, 2-mercaptoethanol and cysteine were compared with CoM with respect to serving as substrates for alkyl transfer reaction by EaCoMT, the enzyme show very low levels of activities and high km indicating low affinities to these thiols and was highly specific for CoM. This could be explained by the interaction of CoM with the strategically placed positively charged residues that interact with them. For effective zinc mediated activation of the thiol group of CoM to take place, the electrostatic interaction between the thiol and the zinc is critical. This could be brought about by holding the rest of the substrate i.e. the sulfonate moiety, in place by strong electrostatic interactions with the enzyme thereby avoiding translation of the substrate in the substrate binding site which could lead to weak interaction between the thiol and the zinc. The other thiols mentioned above lack the crucial determinant, the sulfonate group and hence could be weakly bound to the enzyme only via electrostatic interaction of the thiol with zinc and therefore may contribute to low affinities and very low activities.

Although the EaCoMTs discussed above belong to different genera of microorganisms, their sequences are highly conserved. Analyses of the

homology models have not only provided insights into the alkyl transfer reaction and substrate binding but also have underscored the efficacy of the sulfonate group of CoM to act as a handle in orienting the substrates correctly for catalysis. Firstly, the homology models have revealed that the glutamate residue is highly likely to be the exchangeable ligand of the active site zinc. This is the first time for these enzymes where a new ligand other than the cystienes and histidine is implicated in zinc binding. Mutagenesis studies on the Glutamate residue could confirm its role in zinc coordination and catalysis. Secondly, the above studies have shed light on the putative substrate binding sites in the EaCoMTs. It is very interesting that phylogenetically diverse groups of eubacteria utilize the same cofactor for alkene metabolism and that they have a structurally conserved substrate binding site to bring about a common reaction. The substrate binding pockets in these enzymes which feature the conserved arginine and lysine residues promote strong electrostatic interactions with the sulfonate group of CoM. The arginine and lysine residues are placed strategically with respect to their distance and the way they approach the sulfonate to accommodate and orient it for proper catalysis. The fact that there is a specific structural pattern in the substrate binding sites of these enzymes suggests that they are designed for a specific function common to all these enzymes: utilize the sulfonate group of CoM as a handle in key electrostatic interactions with the conserved residues to assure proper binding and orientation of the substrate for zinc-mediated thiol activation. The reaction carried out by

the EaCoMTs also ensures that the resultant alkyl groups have a key chemical group, the sulfonate moiety of CoM conjugated to them. This is of great importance because, the alkyl groups themselves lack any special chemical groups which could aid them in specific binding to the enzymes those utilize them as substrates.

R- and S-Hydroxypropyl CoM Dehydrogenases: the Role of CoM in Stereo Selective Catalysis

R- and *S*- hydroxypropyl dehydrogenases catalyze the NAD^+ dependent oxidation of 2-*R*- and 2-*S*-hydroxypropyl CoM respectively to the achiral product 2-ketopropyl CoM (2-KPC). These enzymes belong to the short chain dehydrogenase reductase (SDR) family of enzymes, and have the classic Serine, Tyrosine and Lysine catalytic triad [52, 55]. The dehydrogenase pair of enzymes is the only other pair of stereo selective enzymes known to act on opposite enantiomeric substrates in a concerted way in a common pathway. Till their discovery the other pair known to be acting in concert in a single pathway were the tropinone reductases I and II from the plant *Datura stramonium* where these enzymes catalyze the oxidation of the 3-keto group of tropine to tropinone and pseudotropine, respectively. A model was proposed to explain how the two enzymes might be able to bring about stereo selective catalysis. According to this model, differentially placed positively charged residues in the substrate binding sites of these enzymes could provide the switch necessary for

positioning and orienting the substrate in a correct fashion with respect to the active site residues.

The mechanism of stereo selective catalysis was not clear until very recently when crystallographic structure of *R*-HPCDH with product bound at the active site was determined [96]. This structure has provided a structural basis for stereo selectivity in these two dehydrogenases. The structure of *R*-HPCDH suggests that the electrostatic interactions between the sulfonate group and the positively charged residues in the substrate binding site are the key for catalysis. Figure 43 A shows the amino acid environment of the bound product 2-ketopropyl CoM at the substrate binding of *R*-HPCDH. The major interactions of the substrate with the enzymes are the electrostatic interactions of the sulfonate group with the side chains of Arg196 and Arg152. Trp195 acts as a back stop to prevent translation of the substrate during catalysis. A very interesting feature of the substrate binding pocket is the presence of two methionines flanking the substrate binding pocket from two sides. These methionines appear to be shielding the substrate binding site from surrounding solvent molecules that could possibly be present. Also the structure suggests that, after the substrate is bound, these methionines could be shielding the electrostatic interactions of the sulfonate group with the positively charged arginines and the electrostatic interactions of the hydroxyl group of the substrate with the active site residues from unwanted interactions with the solvent. In order to provide the structural basis for stereoselectivity in the two dehydrogenases, a homology model of S-

HPCDH was constructed. Comparison of the crystal structure of *R*-HPCDH and the homology model of *S*-hydroxypropyl CoM dehydrogenase *S*-HPCDH clearly shows that although the overall architecture of both the enzymes are very similar, clear differences can be seen in the substrate binding region (Figure 43 B). The structure of *R*-HPCDH along with the homology model of *S*-HPCDH has not only provided a structural basis for stereoselectivity but also has revealed key substrate-enzyme interactions especially in context to the sulfonate binding site. Although residues in the active site are superimposable, the Arg residues (Arg152 and Arg196) that were observed to be crucial in binding the sulfonate in *R*-HPCDH have been replaced in *S*-HPCDH by Met147 and Gln191, respectively. These side chains would not support the binding of substrate in the same manner as observed for *R*-HPCDH. The absence of the sulfonate binding site at that region clearly suggests that the *S*-HPC substrate cannot be bound in the similar binding mode as the *R*-HPC bound to the *R*-HPCDH. Effective hydride abstraction by short chain dehydrogenases is dependent on the correct orientation of the hydride with respect to the nicotinamide and the correct orientation of the substrate hydroxyl group with respect to the Tyr of the catalytic triad. Superimposition of *S*-HPC bound to *R*-HPCDH on the NAD^+ bound structure of *S*- HPCDH followed by rigid body reorientation so that the hydroxyl and hydrogen groups with the active site residues and NAD^+ puts the sulfonate moiety within hydrogen bonding distance to two positively charged residues

(Arg205 and Lys208). As in the case of *R*-HPCDH, substrate binding is characterized by two positively charged

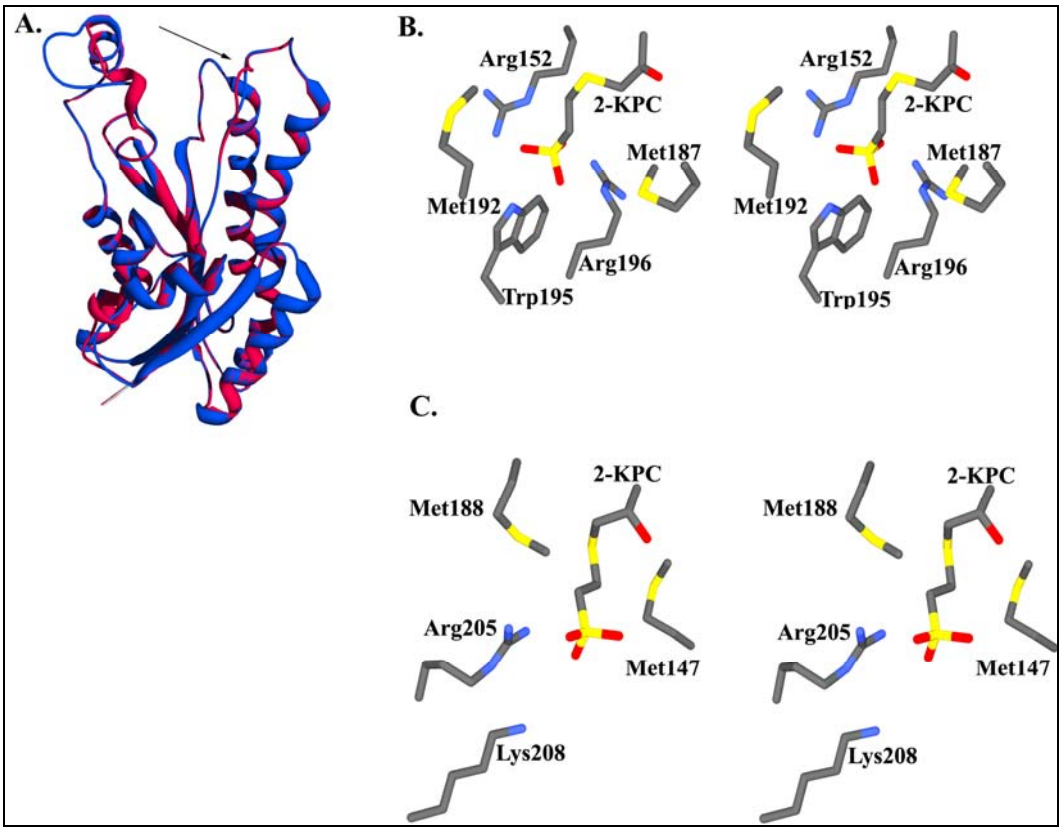

Figure 43. (A). Superimposition of the structure of *R*-HPCDH and a homology model of *S*-HPCDH based on the *R*-HPCDH structure. Arrow indicates the differences in the substarte binding site. Key residues in the substrate binding sites found in (B) the structure of *R*-HPCDH (2CFC.pdb) and implicated in (C) the *S*-HPCDH homology model.

sulfonate binding groups shielded by two flanking methionine residues (Met147 and Met148) (Figure 43 C). The sulfonate binding sites in the *R*- and *S*-HPCDH represent analogous sites involving a different set of amino acid residues at spatially different positions on the respective enzymes. Comparison of the spatial orientation of this putative sulfonate binding site to the sulfonate binding site of *R*-HPCDH to envision the differential placement of positively charged

residues that bind the sulfonate moiety of the substrate has been carried out by superimposing the *S*- and *R*-HPCDH enzymes. Superimposition of the *R* and *S*–HPCDH clearly indicates the differential spatial orientation of the sulfonate binding sites in these two enzymes. From the Figure 44, it is clear that although the orientation of active site residues and the NAD^+ are exactly same, the sulfonate and the methyl groups are switched in order to position the hydrogen and hydroxyl groups of the substrate for hydride and proton abstraction respectively.

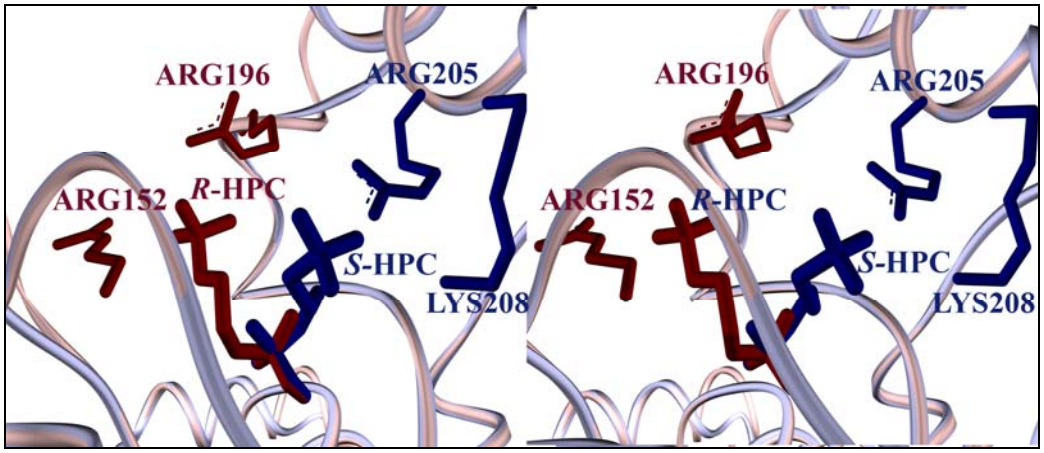

Figure 44. Wall eyed stereo picture of superimposed *R* and S-HPCDH to show the differential spatial orientation of sulfonate binding sites. The *R*-HPCDH residues and *R*-HPC are colored dark pink while the *S*-HPCDH residues and *S*-HPC are colored blue.

The effectiveness of CoM, when conjugated to small substrate molecules which lack innate chemical groups, as a chemical handle is underscored in these structures. *R*- and *S*- hydroxy propylene by themselves would be very small and lack any chemical group on them that could aid in specific binding to the enzymes. Because of their small size, it would be very difficult to design an

active site capable of distinguishing these two substrates which only differ in their chirality. But, conjugation of *R*- and *S*- epoxypropane to CoM in the previous step of the pathway to produce *R*- and *S*- hydroxypropyl CoM has provided these alcohols a handle which could be used for specific binding to the respective enzymes thereby bringing about stereo specific catalysis. By incorporating a common cofactor binding motif in their substrate binding site, these enzymes have effectively exploited the features of CoM to discriminate between enantiomers.

2-KPCC – The Role of CoM in Aligning Substrate for Electron Transfer and Thiol Dependent Reductive Cleavage

The final reaction in the pathway of epoxypropane ring opening and carboxylation is the CO_2 dependent cleavage of 2-KPC to produce acetoacetate and CoM [38]. The 2-KPCC is a homodimeric protein composed of subunits of molecular weight 57 kDa containing one molecule of FAD per subunit. The amino acid sequence reveals the enzyme belongs to the family of DSOR enzymes, which catalyze diverse reactions involving the two electron reduction of a substrate. The enzymes have a redox active cysteine pair which participates in the reduction of the substrate molecule. The reaction catalyzed by 2-KPCC is initiated when NADPH reduces the enzyme flavin, which then reduces the oxidized cysteine pair (Cys82 and Cys87). The substrate 2-KPC then binds, followed by nucleophilic attack of the interchange cysteine thiol on the substrate thioether sulfur, resulting in heterolytic cleavage of the S-C bond. This step

represents a fundamentally novel reaction with respect to the DSOR family, in that a thioether is attacked by cysteine rather than a disulfide. Except for mercuric reductase all the other known members of DSOR family of enzymes attack a disulfide bond. The cleavage of thioether results in the formation of a stabilized carbanion of acetone and a mixed disulfide of cysteine and CoM. Enolacetone then serves as a nucleophile for attack on CO_2 in the carboxylation step. In the final step CoM is regenerated by the reduction of the mixed disulfide of CoM and the interchange thiol.

The crystal structure of 2-KPCC has been solved [66] and it resembles the members of the disulfide oxidoreductase (DSOR) family of enzymes. [39, 111]. The comparison of the substrate bound and native structures suggests that upon substrate binding, the enzyme undergoes a conformational change in which the substrate is encapsulated such that the ketopropyl moiety is located in a hydrophobic environment which excludes the bulk solvent. This sort of encapsulation will promote the formation of the product acetoacetate over protonation and formation of a metabolically unproductive product acetone. The structure of the substrate bound state clearly indicates that substrate is bound at the active site through mainly the strong interactions of two Arg side chains (Arg56 and Arg365) of 2-KPC with the sulfonate group of the substrate. The side chains of Arg 56 and Arg365 approach from the sides and interact directly with CoM sulfonate group by both side chain and main chain interactions. The structure also revealed that the formation and stabilization on an enolate

intermediate could be facilitated by the interaction of a His oriented water molecule with the oxo group of the substrate. Recently the structure of the mixed disulfide state and a CoM disulfide bound state of 2-KPCC has been captured. The mixed disulfide state of the enzyme reveals that the CoM sulfonate binds in the same manner as in the substrate bound state via electrostatic interactions with the arginine residue and a well ordered water molecule is located within hydrogen bonding distance to the disulfide suggesting that it could be important in the protonation of the mixed disulfide upon product reduction and release. Density adjacent to the mixed disulfide is consistent with either acetone (the product) or acetate (a component of the crystallization buffer) and reveals an anion binding pocket that could be important in stabilizing the developing charge during the formation of acetoacetate in the subsequent carboxylation reaction. The CoM disulfide bound structure reveals that one of the sulfonate is bound as revealed in the other structures while the other sulfonate is bound by the aforementioned anion pocket fortifying the suggested role of that pocket in stabilization of the charge on acetoacetate.

Of high relevance to the present study is the interaction of the bound substrate with the enzyme in context with proper orienting of the substrate for thio-ether bond cleavage (Figure 45). It can be seen that the electrostatic interactions between the sulfonate of the CoM and the arginine residues are the key in binding of the substrate. The substrate binding pocket of 2-KPCC is strikingly similar to that of *R*-HPCDH consisting of two arginine residues

approaching from two sides. Further, in 2-KPCC a phenylalanine side chain (Phe57) acts as a backstop preventing further translation of the substrate which is analogous to the role played by Trp195 in *R*-HPCDH. Also, there are two methionines (Met140 and Met361) found flanking the substrate and thereby shielding the electrostatic interactions between the substarte and the enzyme from the surrounding environment.

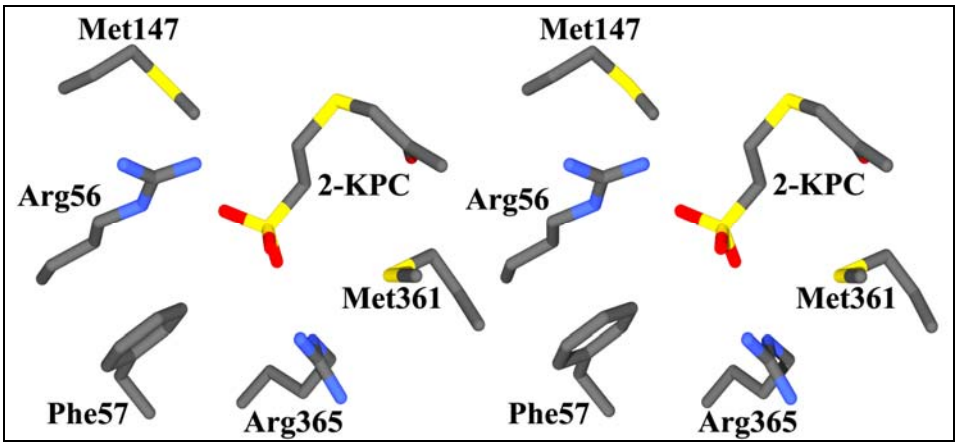

Figure 45. Structural representation of the substrate binding site of 2-KPCC showing the bound substrate 2-KPC (1MO9.pdb). All figures were prepared using Swiss PDB viewer [90] and rendered using Povray [91].

The compact nature of the substrate binding pocket of 2-KPCC reflects on the role of CoM in aligning substrate for electron transfer and thiol dependent reductive cleavage The sulfonate moiety again, in this context, is used as a handle to properly orient the business end of the molecule towards the active site cystienes for catalysis. It should be noted that proper alignment of the thio-ether bond of 2-KPC with respect to the interchangeable cysteine is the key to cleavage and reduction of the substrate. This is brought about by positioning of

key residues in close proximity to the active site such that they can be involved in electrostatic interactions with the sulfonate group of the substrates.

Summary and Conclusions

Specific binding and orientation of short chain alkenes and their derivatives at the active site of the enzymes of microorganisms which metabolize them is a challenge due to the lack of innate chemical groups in these compounds. The above studies suggest that some microorganisms have circumvented this problem by conjugating these reactants with a small, but unique cofactor, CoM whose use was previously thought to be limited to methanogenesis in methanogenic archea. The reactants are conjugated to CoM such that the resulting products posses the sulfonate group of the CoM which could be utilized as a handle to bind and properly orient these substrates at the active site of the enzymes. The role of CoM is discussed in three different contexts. In EaCoMTs the sulfonate group of the CoM is utilized to align the thiol group of CoM for nucleophilic activation of the thiol by zinc. This is brought about by a common CoM/sulfonate binding structural motif involving arginine and lysine residues which are conserved across many genera of microorganisms suggesting a common CoM binding motif. In the stereo specific dehydrogenases, the substarte binding site is designed in such a way that the sulfonate groups of two similar substrates could be bound at two spatially different sites on two structurally similar enzymes to bring about stereo

selectivity. In the third case, 2-KPCC the CoM binding site has exploited the interactions of the enzyme and sulfonate to align the substarte thio-ether linkage with the active site cysteine. It should be added here that the CoM/sulfonate binding site of EaCoMT are less compact compared to the ones in R-HPCDH, S-HPCDH and 2-KPCC. This could be because of the fact that the latter enzymes catalyze highly specific reactions such as stereo specific catalysis and reductive cleavage/carboxylation as opposed to the relatively simple group transfer reaction catalyzed by the former enzyme.

CHAPTER 5

PROGRESS TOWARD THE STRUCTURE DETERMINATION OF ACETONE CARBOXYLASE

Chapter Abstract

Acetone carboxylase is the enzyme involved in the carboxylation of acetone to acetoacetate allowing the utilization of acetone as the sole source of carbon in certain microorganisms. This enzyme has been extensively characterized in the organism *X. autotrophicus*. Much less is known about the mechanism of this enzyme except for the fact that it requires ATP and manganese for catalysis. In this chapter, significant progress has been made towards the determination of the three dimensional structure of acetone carboxylase. Exhaustive trials of optimization of conditions for crystal growth to obtain better quality crystals, heavy atom derivatization and data collection have been carried out. Initial electron density maps calculated from the available phase information derived from a number of moderate quality heavy atom derivatives by the method of MIRAS have revealed a clear protein-solvent boundary. This work has provided the ground work required for further efforts towards successfully determining the structure. The data sets and results available currently will provide a frame work where the heavy atom binding sites obtained from these data sets could be used along with newly found sites to

provide a better structure solution where protein secondary structures can be identified thereby shedding light into the catalytic mechanism.

Introduction

Acetone is an intermediate produced by the fermentative metabolism of certain anaerobic bacteria and during starvation in mammals. Accumulation to higher concentrations is harmful for the bacteria [72, 112] and is metabolized further to eliminate it. Many aerobic and anaerobic microorganisms that can utilize acetone as the sole source of carbon have been identified and isolated. Studies have revealed that acetone metabolism in these bacteria can proceed via two distinct pathways. For some aerobic bacteria acetone is proposed to undergo oxygenation by O_2 dependent monoxygenases to form acetol (hydroxyacetone) as the initial product [112, 113]. For all other bacteria including anaerobes, acetone is found to be carboxylated to acetoacetate or an acetoacetyl derivative as the final product [114, 115]. Very little is known about the enzymes which catalyze the reactions involving the conversion of acetone to different products.

Recently acetone metabolism is studied in depth in the organism *Xanthobacter autotrophicus* capable of aerobic growth utilizing acetone as sole source of carbon [74]. *X. autotrophicus* can utilize either isopropanol or acetone as growth substrates. Isopropanol is converted to acetone by the activity of a secondary alcohol dehydrogenase. Acetone, either as a growth substrate or produced by the activity of alcohol dehydrogenase, is converted directly to the

central metabolite acetoacetate through the activity of an acetone carboxylase (AC). The first evidence that ATP is involved in acetone carboxylation came from the fact that in this organism, acetoacetate production could be reconstituted by addition of ATP in cell extracts [75].

Acetone carboxylase (AC) has been isolated, purified and characterized from *X. autotrophicus* [74, 77]. AC is a heterohexameric protein comprised of three polypeptides with molecular weights of 85,300, 78,300, and 19,600 daltons arranged in a $\alpha_2\beta_2\gamma_2$ quaternary structure [77]. Recent studies have shown that manganese is required for catalysis and is tightly bound to the enzyme [50]. It has also been shown that magnesium added exogenously can enhance the rate of the reaction. This might be because of the formation of MgATP complex which is required for the reaction.

Reaction Mechanism of Acetone Carboxylase

Organic carboxylases in general, use general acid/base chemistry to generate a carbanionic species for attack on CO_2 or bicarbonate. Bicarbonate is a poor electrophile, and therefore most carboxylases use biotin to activate bicarbonate for nucleophilic attack in an ATP-consuming reaction. Notable exceptions are Ribulose bisphosphate carboxylase [116] and 2-KPCC [38], which use low-abundant CO_2 as a direct electrophile [116], and phosphoenol pyruvate carboxylase (PEP-C) [17], which activates bicarbonate to electrophilic carboxy phosphate using the enol phosphate of Phosphoenol pyruvate (PEP). AC is

unusual amongst the carboxylases since catalysis requires the hydrolysis of ATP to 1 mol AMP and 2 mol of inorganic phosphate [14, 74, 75]. Biotin is not associated with acetone carboxylase and the pKa of the methyl proton of acetone that must be abstracted is very high (~18). ATP hydrolysis in AC may have two functions: to allow the formation/stabilization of enolacetone, and to activate bicarbonate as an electrophile. Acetone carboxylation resembles the combination of two reactions that when added together achieve pyruvate carboxylation: PEP synthetase [117-120] and PEP carboxylase. It is unknown how the hydrolysis of two high energy bonds of PEP synthetase facilitates the formation of PEP but it has been suggested that the reaction proceeds via a phosphoenzyme intermediate [117] (Figure 46 A). Recent studies in the Ensign group involving isotopic exchange indicate that alpha-proton abstraction from acetone occurs in concert with transfer of the gamma-phosphoryl group of ATP to the carbonyl oxygen, generating phosphoenol acetone as the activated nucleophile for attack on carbon dioxide [121] (Figure 46 B).

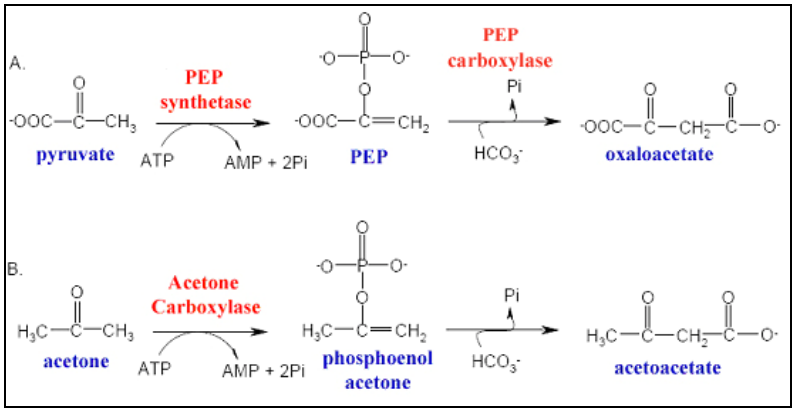

Figure 46. (A) Reaction mechanism of PEP synthetase. (B) Proposed reaction mechanism of AC involving a phosphoenol acetone intermediate.

Determination of the Structure of Acetone Carboxylase

Very little is known about the mechanism of catalysis of either the prokaryotic PEP synthetases [117, 119, 120, 122, 123] or the acetone carboxylases. X-ray crystallography has become an indispensable tool to understand mechanism of enzymes. The three dimensional structure determined by X-ray diffraction methods give a detailed idea of the secondary,tertiary and quaternary structural elements. The product or substrate bound structures can be solved by collecting data on the crystals which have been either co-crystallized or soaked in the product or the substrate. Such structures not only allow us to envision the substrate or product binding at the active site but also allow us to decipher the mechanism of the enzyme catalysis. In many cases where much less is known about the enzyme active site, substrates and products can be soaked into the crystal and when the substrate/product bound structure is observed, one can get an idea about the active site residues. The wide range of advantages that X-ray crystallography can provide could be utilized to study an enzyme like AC. As mentioned earlier, not much is known about the mechanism of AC. The three dimensional structure along with bound substrates like acetone and ATP or the structure where product acetoacetate is bound can shed light on the catalytic process. Since acetone carboxylase does not resemble any other enzymes in the data base, the structure could potentially have unique features with respect to its tertiary structure. Since not much is known about the mechanism of AC, an

imperative must be placed on the determination of the structure of AC to further advance the understanding of its mechanism.

Crystallization of Acetone Carboxylase

Acetone carboxylase was first crystallized using the hanging drop vapour diffusion method with 0.2 M $MgCl_2(6H_2O)$, 0.1 M Tris HCl pH 8.2, 30% PEG 6000, and 20% glycerol as a precipitating solution [124]. The crystals grew after 2-3 weeks of incubation and were colorless and diamond shaped measuring about 0.3x0.3x0.3 mm. These crystals belonged to the space group $P2_12_12_1$ with cell dimensions a=76.2 Å, b=122.3 Å, and c=264.2 Å. This crystallization mother liquor is amenable to direct cryo-cooling and data were collected at beam-line 9-1 of the Stanford Synchrotron Radiation Laboratory (SSRL). The crystals clearly diffracted beyond 2.8 Å resolution and a complete data set to 3.2 Å resolution was collected [124]. Since these crystals were not reproducible, different conditions were screened to obtain crystals. Reasonable quality crystals were obtained by modifying cryo 6 (0.16M magnesium chloride hexahydrate, 0.08 M TRIS, 24% w/v polyethylene glycol 4000 and 20% v/v glycerol) condition from Hampton research [125] where 7% PEG was used instead of 24%. The quality of diffraction of these crystals was improved upon adding hexamine cobalt chloride as an additive during crystallization up to a final concentration of 0.01M. The crystals grew to a good size of 0.4 x 0.5 x 0.5 mm (Figure 47) and belonged to the space group $P2_12_12_1$ with unit cell dimensions a=121.3, b=152.2, c= 264.5.

Matthews' coefficient [126] indicated that a complete hexamer is present in the asymmetric unit which allows the use of non-crystallographic symmetry (NCS) [127] in solving the structure. In the same crystallization condition, acetone carboxylase crystallizes frequently in another space group $P2_1$ with cell parameters a=76.2Å, b=121.3Å, c= 264.2Å and β=91.3°.

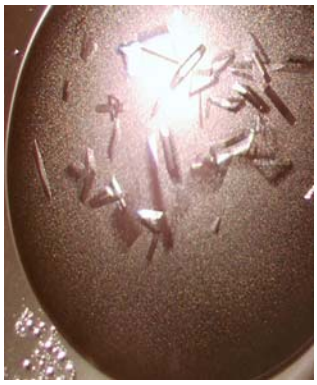

Figure 47. Crystals of acetone carboxylase.

One of the methods for structure determination is molecular replacement (MR). If a related or similar protein structure is already known then the method of MR can be used [128]. In MR, rotation and translation functions are carried out to find the rotation and translation operations which position the search model structure in the unit cell of the unknown crystal structure so as to give the highest correlation between experimental diffraction measurements and those calculated from the model. This method relies on the existence of a known related structure and since AC does not have sequence similarities to any of the known protein structures so far solved, MR cannot be yet used to solve the structure of AC. With molecular replacement not an option, some of the other

methods to obtain structure solution need to be exploited. Some of the other routinely used methods are MAD [129] and Multiple isomorphous replacement with anomalous scattering (MIRAS) [130]. Multiple wavelength anomalous dispersion is an increasingly popular method used to obtain structure solution. This method exploits the anomalous scattering properties of certain elements. X-rays set the electrons in an atom vibrating and these vibrations generate radiation of the same frequency, which is propagated in all directions. This is the normal scattering which gives rise to the diffraction. The diffracted radiation is in phase with the incident radiation. If however, the energy of the incident radiation is close to the transition energy which can excite the electron, the diffracted radiation has a different phase. Also the intensity of the diffracted radiation is also decreased because some of the energy is absorbed to bring about the excitation of electrons. This is called anomalous scattering. At wavelengths of X-rays convenient for diffraction experiments (less than 1.6 Å), atoms lighter than phosphorous or sulfur behave as normal scaterrers because they have no transitions at the corresponding energy of this wavelength. Anomalous scatterers include elements from iron (molecular mass 26) to palladium (molecular mass 46) at K absorption edges and much heavier atoms like lanthanides and beyond give strong anomalous effects from the L edges. Tunable synchrotron wavelengths make this method more powerful because measurements can be made at a precisely chosen wavelength. In the method of MAD, phase information to solve the structure is obtained by measuring the scattering of X-

rays by the anomalous scatterer along with the protein. Anomalous scatterers could be intrinsically present in the protein in question or could be introduced into the crystals either by co-crystallization or by soaking the crystals into solutions of salts of anomalous scaterrers. AC, as mentioned earlier, has two manganese ions tightly bound to it. This was verified by isothermal titration calorimetry experiments. Manganese is capable of anomalous dispersion at wavelengths 1.8Å and has been used in solving the structure of many proteins. For acetone carboxylase Matthews' coefficient suggests that one hexamer is present in an asymmetric unit which makes the number of manganese atoms in the asymmetric unit to be also two. Compared to the number of amino acid residues in the asymmetric unit, the number of manganese atoms is too few to produce an anomalous signal sufficient to give enough phase information to solve the structure. Another method to incorporate anomalous scaterrers into the protein is to express the protein heterologously in *Escherichia coli* cells in the presence of selenomethionine so that the selenium gets incorporated into the protein. This has not yet been possible in AC since numerous trials to successfully express measurable quantities of AC in *E.coli* have been unsuccessful. Recently, the anomalous scattering property of sulfur has been used to solve structures of many proteins. Since proteins frequently contain number of sulfur atoms in cystiene and methionine residues, there is no need to incorporate any other anomalous scaterrers. Unfortunately for AC, there are too many sulfur atoms (120) in the asymmetric unit and location of these atoms can

be challenging using traditional Patterson methods especially if the number of heavy atoms is large. This method has been tried for acetone carboxylase without any success. The positions of sulfur atoms were not found in the difference Patterson maps using the program SOLVE. With selenium- and sulfur- MAD currently not feasible, anomalous scatterers other than these can be introduced into the crystals and can be exploited for MAD experiments. This is done by soaking the crystals in solutions of salts containing the anomalous scatterers.

Another method of obtaining structure solution is multiple isomorphous replacement (MIR) [130]. In this method heavy atoms (atoms with high atomic numbers) are introduced into the crystal lattice and diffraction data is collected both on the crystals which are free of heavy atoms (generally termed “native”) and on the crystals which contain heavy atoms (termed “derivative”). This technique is based on the fact that due to the perturbations caused by the introduction of heavy atoms, in the amplitudes of the diffracted waves, some amount of phase information can be derived. MIR method entails that the crystals are isomorphous with the native crystals after soaking in heavy atom solutions. This means that there is no considerable amount of change in the cell parameters of the crystal after it has been subjected to heavy atom soaking. A single derivative is not enough to solve the phase problem and obtain a structure solution. MIR requires data from at least two derivatives along with the native data set. Recently, a combination of anomalous dispersion with multiple

isomorphous replacement has been used routinely to obtain structure solutions of many proteins. This method is termed multiple isomorphous replacement with anomalous scattering (MIRAS). In MIRAS several heavy atoms which are also anomalous scaterrers at a particular wavelength of X-rays are substituted isomorphously into a protein and the differences occurring due to both isomorphous replacement and anomalous dispersion in amplitudes are measured. This additional observation improves the accuracy of phase determination.

Heavy Atom Derivatization of Acetone Carboxylase Crystals

Two different approaches were taken to produce heavy atom derivatives or to introduce anomalous scatterers into the crystals of acetone carboxylase. The first method was to grow the crystals in presence of heavy atom derivatives. Second method was to soak the crystals in heavy atom solution. Both these methods have been successful for obtaining crystals suitable for structure determination using MIR and MAD in our laboratory [66, 96].

Co-crystallization of AC with Heavy Atoms

The heavy atoms used in co-crystallization trials of AC are listed in Table 9. A 20 mM solution of different heavy atoms were prepared in water and added to the protein solution (45mg/mL) to give final concentrations of 1mM, 2 mM, 4 mM, 5mM, 7 mM and 10 mM respectively. Different concentrations of heavy atoms were used to access the tolerance of the protein solution towards precipitation. Proteins generally tend to precipitate upon contact with heavy

atom solutions. The concentrations which precipitated the proteins (7 mM and 10 mM) were not used. The other concentrations tested did not precipitate the protein and were used in the co-crystallization trials. The protein along with the heavy atoms was incubated at room temperature for half an hour to allow incorporation of heavy atoms. It was observed that crystals grew only in presence of mercury salts, namely Mercury (II) chloride, Ethylmercurithiosalicylic acid, sodium salt, Mercury (II) potassium iodide, Ethylmercury chloride and Mercury (II) iodide at concentrations 2mM, 4mM and 5mM respectively. The crystals belonged to the space group $P2_1$ with cell dimensions. a=76.4 Å, b= 121.2 Å, c=264.3 Å and β=91.2°. The solvent content analysis of the crystals was 60% [126] indicated that a complete hexamer is present in the asymmetric unit allowing the use of NCS in structure determination. Crystals failed to grow in other heavy atom salts and resulted in precipitation of protein. Data sets were collected on one crystal each which grew in the presence of mercury(II) chloride and EMTS. The other crystals had poor quality diffraction and were not considered for data collection.

Soaking the Crystals in Heavy Atom Solutions

To begin with, an initial screening for suitable heavy atoms was carried out. The heavy atom salts used for derivatization are listed in Table 9. A 20 mM solution of each heavy atom was prepared in the mother liquor which was used to grow the crystals and was added to the hanging drop to give a final concentration of 0.5 mM,1 mM,2 mM, 5 mM, 7 mM and 10 mM. The crystals were left in the

heavy atom for varying amounts of time (0.5, 1, 5, 12, 24 and 48 hours). The combination of heavy atom concentration and times of soak which resulted in crystals which are capable of good diffraction was found out be 5 mM for 12 hours and thus selected for data collection. The uptake of the heavy atom by the crystals is indicated by the change in color of the crystal if the heavy atom compound is colored. The crystals were cryo-cooled using liquid nitrogen. Soaking the crystals in heavy atom solutions using the above mentioned method suggested that the introduction of heavy atom into the crystals might have caused perturbations in the unit cell parameters and might have rendered them non isomorphous. This was due to the fact that crystals, when soaked, appeared to undergo morphological changes such as cracking or became very brittle. The problem of cracking was minimized by slow dialysis of the heavy atom solution into the crystals which were placed in dialysis buttons rather than directly adding the heavy atom solution to the drop containing the crystals as mentioned earlier. Crystals soaked in a number of heavy atom solutions were not capable of diffraction. The crystals which diffracted after the heavy atom soaks were considered for further data collection The data collection statistics reported here are for the crystals which diffracted after the heavy atom soak and the names of the heavy atoms are mentioned for each set of data.

Table 9. Heavy atom salts used in the derivatization of AC.

1. Gold (I) Potassium Cyanide
2. Gold (III) Chloride
3. Potassium Tetrachloroaurate
4. Mercury (II) chloride
5. Mercury (II) acetate
6. Mercury (II) bromide
7. Mercury (II) iodide
8. Mercury (II) cyanide
9. Mercury (II) oxide
10. Tetrakis(acetoxymercuri)methane
11. Methylmercury (II) chloride
12. Ehtylmersury thio salicylate (EMTS)
13. Thallium (I) Chloride
14. Thallium (III) Acetate hydrate
15. Lead (II) Acetate trihydrate
16. Lead (II) Nitrate
17. Lead (II) Chloride
18. Triethyl Lead Acetate
19. Silver Nitrate
20. Cadmium Chloride hydrate
21. Cadmium Iodide
22. Potassium Hexachloroiridate (IV)
23. Iridium (III) Chloride hydrate
24. Sodium Hexachloroiridate (III) hydrate
25. Ammonium Hexachloroiridate (III) hydrate
26. Potassium hexanitroiridium (III)
27. Potassium Osmate (IV) dehydrate
28. Ammonium Hexabromoosmate (IV)
29. Potassium Hexachloroosmate (IV)
30. Osmium (III) Chloride hydrate
31. Sodium Tungstate dehydrate
32. Ammonium Tetrathiotungstate (VI)
33. Samarium (III) Chloride hexahydrate
34. Samarium (III) Acetate hydrate
35. Samarium (III) Nitrate hexahydrate
36. Lanthanum (III) Nitrate hexahydrate
37. Europium (III) Nitrate hexahydrate
38. Europium (III) Chloride hexahydrate
39. Gadolinium (III) Chloride hydrate
40. Lutetium (III) Chloride hexahydrate
41. Lutetium (III) Acetate hydrate
42. Ytterbium (III) Chloride hydrate

Table 9. Heavy atom salts used in the derivatization of AC-continued.

43. Dysprosium (III) Chloride hexahydrate
44. Praseodymium (III) Chloride heptahydrate
45. Neodymium Chloride hydrate
46. Holmium (III) Chloride hexahydrate
47. Potassium Hexachlorohenate (IV)
48. Potassium Perrhenate
49. Potassium Tetrachloroplatinate (II)
50. Ammonium Tetrachloroplatinate (II)
51. Potassium Hexachloroplatinate (IV)
52. Potassium Tetranitroplatinate (II)
53. Potassium Tetracyanoplatinate (II)
54. Dichloroethylenediamine Platinum (II)
55. Diammino Platinum Dinitrate
56. Potassium Tetrabromoplatinate (II)
57. Potassium Hexabromoplatinate (IV)
58. Platinum Potassium Iodide
59. Platinum Potassium Thiocyanate
60. Di-μ-iodobis(ethylenediamine)diplatinum (II) Nitrate
61. cadmium chloride

Use of Heavy Atom Clusters/Complexes to
Derivatize Crystals

One of the problems of soaking heavy atoms into the crystals is the incorporation of too many heavy atoms into the asymmetric unit. This can happen due to the non specific binding of heavy atoms to the protein. Incorporation of too many sites makes it problematic to find their positions. There are ways to circumvent this problem like back-soaking and use of heavy atom clusters. Back-soaking involves soaking the crystal in heavy-atom solution for a known time and then removing it back into its heavy atom-free stabilizing buffer for anything from a few minutes to a period of days. The purpose of this procedure is to remove any heavy atoms which are non-specifically bound and also to reduce the

likelihood of additional partially occupied sites caused by weak heavy-atom binding. The other method which is the use of heavy atom clusters will minimize non specific binding because of their huge size. Both these methods have been tried on the AC crystals.

It has been previously demonstrated that the structures of big proteins can be solved by derivatizing them with metal cluster complexes rather than simple heavy atom salts [131]. Such heavy atom clusters possess more than one copy of the anomalous scatterer and hence are capable of producing a stronger anomalous signal and hence make significant contributions to obtaining phase information. One such complex is the tantalum bromide cluster ($[Ta_{(6)}Br_{(12)}]\ Br_2\ 6H_2O$). The $[Ta_{(6)}Br_{(12)}]\ Br_2\ 6H_2O$ cluster is known to be a powerful reagent for derivatization of crystals of large macromolecules at low resolution [132, 133]. The structure of tantalum bromide cluster is shown in Figure 48. The cluster is a regular octahedron of six Ta atoms with 12 bridging Br atoms at the edges of the octahedron.

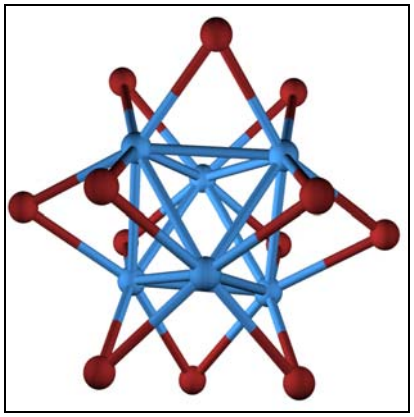

Figure 48. Tantalum bromide cluster.

The cluster is compact, of approximately spherical shape, with a radius of about 6 Å. Both tantalum and bromine display a significant anomalous diffraction signal at their absorption edges at 1.25 Å and 0.92 Å, respectively. Ta_6Br_{12} has been applied so far to solve the structures of the large protein assemblies of the proteasome (673 kDa) [134], transketolase (680 residues per asymmetric unit) [135], 1992), ribulose-1,5-bisphosphate carboxylase/oxygenase (550 kDa) [136-138], and cyclohydrolase I (homodecamer 250 kDa) [139]. Since the heavy atoms are clustered together in big clusters like tantalum bromide cluster, the anomalous signal and isomorphous differences are more and hence the phasing power of these derivatives is higher.

A solution of tantalum bromide cluster was prepared by dissolving the cluster in distilled water to give a final concentration of 10 mM. To each drop containing the crystal, 1 microliter was added at the edge of the drop and allowed to diffuse over 5 hours. After 5 hours another microliter of the cluster solution was added. The procedure was repeated once more after 5 hours. This gave a final concentration of 5 mM in the drops. The crystals appear bright green after soaking in the tantalum bromide solution suggesting uptake of the cluster by the crystals (Figure 49). It was seen that the crystals did not suffer any visible cracking or deformation due to the soaking of tantalum bromide cluster. Although the crystals did not suffer any visible physical damage, the dialysis method of soaking was also carried out for soaking the tantalum bromide cluster into the crystals.

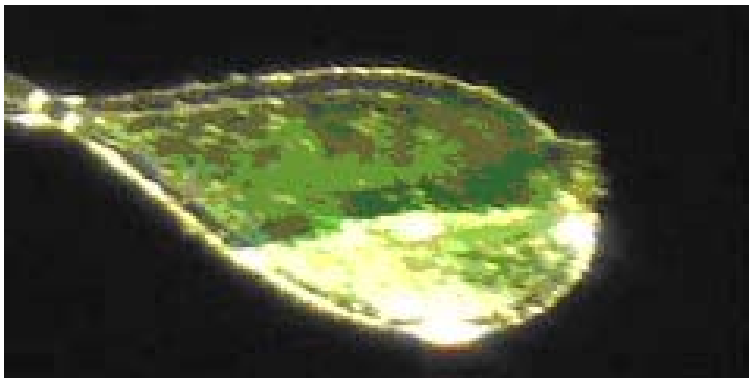

Figure 49. Acetone carboxylase crystal soaked in tantalum bromide and mounted on a rayon loop and flash frozen in liquid nitrogen.

The crystals were harvested at varying intervals of time (1, 2, 3,4,7,9 and 12 hours after the addition of the last aliquot of tantalum bromide cluster solution), flash frozen in liquid nitrogen and tested to see if they still retained the ability to diffract. The crystals retained the ability to diffract even after 12 hours and the soaking time of 12 hours was made as standard for tantalum bromide soaking. Concentrations greater than 5 mM were tried in the soaking experiments and were found (7 mM, 8 mM and 10 mM), to disrupt the crystals by forming cracks and hence these concentrations were not considered for data collection. Co-crystallization of AC with tantalum bromide was also tried at different concentrations of tantalum bromide cluster (1 mM 2 mM and 3 mM), but resulted in protein precipitationat all concentrations within hours of crystallization setup. A three wavelength MAD data set was collected on tantalum bromide cluster soaked crystals. The data statistics for the MAD data set are reported later in this chapter.

Data Collection for Multiple Wavelength Anomalous Diffraction Experiments

The diffraction quality of acetone carboxylase crystals is not very good (4-5 Å) but diffraction to higher resolutions (better than 2.8 Å) can be achieved with X-rays of higher flux such as those from a synchrotron source. Data sets were collected on both native and heavy atom soaked crystals of acetone carboxylase at beam line 9-1, 9-2, 1-5 and 11-1 at SSRL using X-rays at 1Å wavwlength (12800eV). Since acetone carboxylase crystallizes with large unit cell dimensions, care was taken during data collection to ensure good spot separation by adjusting the crystal to detector distance. A distance of 400 mm on the MAR 325 CCD detector at that particular energy gave good spot separation. Because the wavelengths of X-rays are tunable at a synchrotron source, the method of MAD can be exploited to the fullest extent at these sources. The synchrotron source allows precise measurements of diffraction data at different wavelengths. Moreover, since the data is collected on a single crystal the problem of non-isomorphism is eliminated. AC crystals subjected to heavy atom soaks were subjected to excitation scans at particular wavelengths corresponding to the absorption edge of the heavy atom present in the crystals. Once the peak and inflection energies were obtained, complete data sets were collected at those particular energies. A complete data set was collected at energy remote from the peak and inflection as well except for krypton derivatized crystals since those

crystals lost their ability to diffract after data collection at peak and inflection energies. Data sets were processed using HKL2000 [81].

MAD Phasing with Tantalum Bromide Cluster

Crystals were soaked in tantalum bromide cluster as explained earlier. A fluorescence scan was carried out at the synchrotron source to identify the wavelengths corresponding to peak, remote and inflection energies for tantalum and a three wavelength data set was collected and data statistics are reported on Table 10. Heavy atom searches were carried out using the program SOLVE. The heavy atom coordinates, B factors and occupancies are given in Table 11. As it can be seen from the table, the figures of merit at different resolution are lower than 0.20.

Table 10. Multiple wavelength anomalous diffraction data statistics for crystals soaked in tantalum bromide cluster.

Space group	$P2_12_12_1$		
Cell dimensions	a = 122.338 Å b = 154.407 Å c = 265.775 Å $\alpha = \beta = \gamma = 90^o$		
	Peak λ=1.25499 Å	Remote λ= 0.87929 Å	Inflection λ=1.25551 Å
Resolution (Å)	50.0-4.00	50.0-6.00	50.0-6.00
Completeness (%)	95.1 (94.2)	91.1 (97.9)	95.6 (92.5)
Observed reflections	211812	112551	112037
Unique reflections	30638	67068	80914
I/σ	18 [86]	25 [86]	26 [86]
R_{merge}[a]	0.188 (0.314)	0.124 (0.276)	0.134 (0.234)

Numbers in parentheses indicate values for the highest resolution bin.
a. $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle|/\Sigma_{hkl}\Sigma_i|\langle I\rangle|$, where I_i is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.

Table 11. Results of MAD phasing experiment for tantalum bromide derivative of acetone carboxylase.

Resolution range (Å)	FOM	Heavy atom sites					
			X	Y	z	Occupancy	B factor
11.13-8.75	0.29	1	135.6	25.4	31.6	1.49	26.6
8.75 - 7.44	0.27	2	38.7	85.3	28.7	0.91	39.9
7.44 –	0.22	3	137.7	84.7	2.41	0.38	45.0
6.58	0.23	4	39.1	225.0	11.8	0.21	60.0
6.58 –	0.20	5	114.5	6.1	11.4	0.01	60.0
5.97	0.10	6	18.0	19.8	11.9	0.05	55.0
5.97-5.49		7	63.7	46.3	5.8	0.08	30.0
5.49-5.10		8	110.6	224.5	8.8	0.09	45.5

The heavy atom sites were different from the ones got by using the peak wavelength of this data in SIR and MIR phasing experiments. The phases calculated from the MAD experiment were used in computing the electron density map and it was found that these maps were not interpretable.

MAD Phasing with Platinum Potassium Iodide Derivative

Acetone carboxylase crystals were soaked in 5 mM potassium platinum iodide solution by the dialysis procedure explained earlier and a three wavelength data was collected on a single crystal after harvesting and flash freezing in liquid nitrogen. The MAD data statistics are given in Table 12. The data sets were processed using the program HKL2000 and were subjected to heavy atom search using the program SOLVE. The coordinates, occupancies and B factors of the heavy atoms are provided in Table 13. Also the figures of merit at various resolutions are provided in the same table. It can be seen that the

figures of merit are very low at different resolutions. Maps calculated by the using the phases obtained by MAD experiment were not interpretable even after density modification procedures.

Table 12. Multiple wavelength anomalous diffraction data statistics for crystals soaked in Platinum Potassium Iodide.

Space group $P2_12_12_1$

Cell dimensions a = 120.149 Å b = 152.178 Å
c = 263.908 Å $\alpha = \beta = \gamma = 90^o$

	Peak λ=1.0794 Å	Remote λ= 0.89194 Å	Inflection λ=1.07222 Å
Resolution (Å)	50.0-3.50	50.0-4.50	50.0-6.00
Completeness (%)	99.2 (100)	98.8 (99.7)	97.9 (99.3)
Observed reflections	458119	217925	144062
Unique reflections	62235	29928	26438
I/σ	26 (8.0)	24.1 (10)	20.1 (9.5)
R_{merge}[a]	0.116 (0.285)	0.141 (0.200)	0.097 (0.23)

Numbers in parentheses indicate values for the highest resolution bin.
a. $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle|/\Sigma_{hkl}\Sigma_i|\langle I\rangle|$, where I_i is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.

Table 13. Results of MAD phasing experiment for platinum potassium iodide derivative of acetone carboxylase.

Resolution range (Å)	FOM	Heavy atom sites					
			x	Y	z	Occupancy	B factor
11.13-8.75	0.17	1	113.6	55.4	31.6	0.49	56.6
8.75 - 7.44	0.16	2	38.7	65.3	212.	0.51	49.9
7.44 –	0.14	3	137.7	44.5	7	0.78	55.0
6.58	0.13	4	34.1	23.6	33.4	0.01	60.0
6.58 –	0.10				1		
5.97					78.8		
5.97-5.49							

Use of Krypton for Derivatization

Under gas pressures ranging from 8 to 20 bar, krypton is able to bind to discrete sites in hydrophobic cavities, ligand and substrate binding pockets, and into the pore of channel-like structures. These krypton complexes can be used to map hydrophobic sites in proteins, or as heavy-atom derivatives in the isomorphous replacement method of structure determination [140]. Krypton (Kr) also acts as an anomalous scatterer making the crystals amenable for MAD experiments. As the Kr K edge (14.3 keV) is accessible on most MAD beam lines, Kr derivatization provides the additional opportunity to conduct a MAD experiment and obtain phases using only a single crystal. In order to derivatize the acetone carboxylase crystals, they were removed from the mother liquor and immediately placed inside the pressurization cell [141]. A small eppendorf tube containing water had already been placed inside the cell to prevent crystal dehydration. Different amounts of pressures were applied to introduce the krypton gas ranging from 1 MPa to 3 MPa and it was found that the optimum pressure was 2.7 MPa. The cell was pressurized with 2.76 MPa of Kr gas for 1, 2 and 3 minutes, depressurized and quickly opened. The pin was removed using a magnetic wand and the sample was flash frozen using liquid nitrogen. Data sets were collected at two different wavelengths corresponding to peak and remote energies for krypton at SSRL beam line 9-2.

Before the third data set corresponding to the inflection energy could be collected, the crystal lost its capability to diffract. The data sets were processed

using the program HKL2000 and subjected to heavy atom search by the program SOLVE. Data statistics are available upon request. The program unfortunately couldn't find any heavy atom sites and thus this data was not useful in providing any information about heavy atom sites.

It can be seen that none of the MAD experiments using different derivatives of AC such as tantalum bromide cluster, platinum potassium iodide, krypton and mercury provided good phase information. Therefore instead of the MIR approach was taken to see if better phase information would be available from these experiments.

Single Wavelength Data Collection on Native and Derivatized Crystals of AC

As mentioned earlier acetone carboxylase crystallizes in two different space groups namely $P2_1$ and $P2_12_12_1$. Data sets were collected on both these crystal forms and the statistics and the phase information provided by those data sets are discussed in the following sections. The diffraction quality of acetone carboxylase crystals is not very good (4-5 Å) but diffraction to higher resolutions (better than 2.8 Å) can be achieved with X-rays of higher flux such as those from a synchrotron source. Data sets were collected on both native and heavy atom soaked crystals of acetone carboxylase at beam line 9-1, 9-2, 1-5 and 11-1 at SSRL using high energy (12800eV) radiation. Since acetone carboxylase crystallizes with large unit cell dimensions, care was taken during data collection to ensure good spot separation by adjusting the crystal to detector distance. A

distance of 400 mm on the MAR 325 CCD detector at the energies mentioned gave good spot separation.

Data Collection and Experiments on Native and Derivatized Crystals of AC belonging to Space Group $P2_12_12_1$

Tables 14,15 and 16 shows data statistics for native data sets collected on three different crystals, respectively. Tables, 17 and 18 show the data statistics for various heavy atom derivatives of AC. The data sets for the derivatives were collected at different wavelengths which corresponded to the peak energy at which the anomalous scattering properties of the heavy atoms are highest.

Table 14. Data statistics of crystals of AC with no heavy atoms (native set 1).

Wavelength Used (Å)	1.2000
Space Group	$P2_12_12_1$
Cell dimensions	a=121.9, b=152.4 c=264.2, $\alpha = \beta = \gamma = 90^o$
Resolution (Å)	50-3.0
Number of Reflections (number of unique reflections)	223450 (46237)
*R*sym	0.040 (0.15)
% Completeness	99.5(99.1)

Table 15. Data statistics of crystals of AC with no heavy atoms (native set 2).

Wavelength Used (Å)	1.2000
Space Group	$P2_12_12_1$
Cell dimensions	a=117.7, b=152.6, c=263.8 $\alpha = \beta = \gamma = 90^o$
Resolution (Å)	50-3.3
Number of Reflections (number of unique reflections)	237250 (47127)
*R*sym	0.080 (0.10)
% Completeness	98.5(99.1)

Table 16. Data statistics of crystals of AC with no heavy atoms (native set).

Wavelength Used (Å)	0.9000
Space Group	$P2_12_12_1$
Cell dimensions	a=118.7, b=152.4, c=264.4 α = β= γ = 90°
Resolution (Å)	50-3.2
Number of Reflections (number of unique reflections)	227170 (45197)
*R*sym	0.070 (0.088)
% Completeness	97.5(99.1)

Table 17. Data statistics of crystals of AC soaked in various heavy atoms.

Heavy atom	Sodium Tungstate dihydrate	Potassium Hexabromo platinate	Sodium Bromide	Potassium Hexachloro platinate
Wavelength Used (Å)	1.21437	1.0091	0.90605	1.0091
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	A=120.0 b=152.7 c=264.2 α = β= γ = 90°	a=120.4 b=152.1 c=265.7 α = β= γ = 90°	A=120.1 b=153.6 c=263.8 α = β= γ = 90°	a=119.4 b=152.3 c=264.2 α = β= γ = 90°
Resolution (Å)	50-3.10	50-3.5	50-7.00	50-3.7
Number of Reflections (number of unique reflections)	323553 (86911)	323210 (87922)	147079 (19554)	323210 (77910)
*R*sym	0.070 (0.043)	0.090 (0.165)	0.180 (0.296)	0.093 (0.145)
% Completeness	99.9 (97.6)	99.6 (98.7)	99.9 (97.0)	99.3 (98.9)

Table 18. Data statistics of crystals of AC soaked in various heavy atoms.

Heavy atom	Gold chloride	Platinum potassium iodide	Tantalum Bromide cluster	Mercury (II)chloride
Wavelength Used (Å)	0.89674	1.0587	1.25480	1.25480
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)	50-2.7	50-2.9	50-4.0	50-4.0
Cell Dimensions (Å)	a=121.9 b=152.4 c=264.2 α = β= γ = 90°	a=122.0 b=152.7 c=264.2 α = β= γ = 90°	A=119.4 b=152.1 c=265.7 α = β= γ = 90°	a=119.4 b=152.1 c=265.7 α = β= γ = 90°

Table 18. Data statistics of crystals of AC soaked in various heavy atoms-continued.

Number of reflections (number of unique reflections)	487861 (131311)	434156 (117858)	275432 (47802)	275432 (47802)
*R*sym	0.065 (0.220)	0.058 (0.264)	0.06 (0.12)	0.06 (0.12)
% Completeness	99.5 (98.7)	98.4 (97.3)	98 (97.1)	98 (97.1)

Analysis of Native Data Sets

The native data sets were processed using HKL2000. In order to see the degree to which they are different from each other, these data sets were scaled together. The R_{merge} values can give an estimate of non-isomorphism between two crystals. The R_{merge} calculated after scaling of data sets in Table 14 and Table 15 (two native data sets) and listed for different resolution ranges in Table 19. This suggests that these crystals are non-isomorphous but can be treated as isomorphous at lower resolutions. The scaling of the data sets reported from Table 14 and Table 15 gives R_{merge} values which are listed for different resolution in Table 20 which suggests that these data sets do not merge well. One of the reasons may be non-isomorphous nature of the crystals. It can be seen from the Tables 19 and 20 that at lower resolutions the crystals can be treated as isomorphous. This scaling was done to reveal the degree of non – isomorphous nature in these crystals.

Table 19. R_{merge} as a function of resolution for two native data sets (1 and 2).

Resolution bins (Å)	R_{merge}
50.00 6.89	0.070
6.89 5.47	0.085
5.47 4.78	0.090
4.78 4.34	0.102
4.34 4.03	0.110
4.03 3.79	0.132
3.79 3.60	0.140
3.60 3.45	0.190
3.45 3.31	0.220

Table 20. R_{merge} as a function of resolution for two native data sets.(2 and 3).

Resolution bins (Å)	R_{merge}
50.00 6.89	0.060
6.89 5.47	0.075
5.47 4.78	0.090
4.78 4.34	0.112
4.34 4.03	0.120
4.03 3.79	0.152
3.79 3.60	0.170
3.60 3.45	0.180
3.45 3.31	0.230

Heavy Atom Searches, SIR and MIR Phasing.

Each derivative data set was processed using HKL2000 and scaled against the native data set shown in Table 14. This data set was used since the R_{merge} values with other native data sets with the individual derivative data sets were very high. Heavy atom searches were carried out by the program SOLVE [82]. SIR phases were calculated using individual derivatives (Potassium hexabromoplatinate (IV), gold (III) chloride, potassium tetrachloroplatinate, mercury (II) chloride, sodium tungstate dehydrate, sodium bromide

platinum potassium iodide, tantalum bromide cluster) to choose the best set of derivatives for calculation of phases using MIR method. Table 21 shows the results for SIR phasing using different derivatives of AC. The table shows the coordinates for the heavy atoms, their occupancies and B factors. The heavy atom derivatives are scaled against the native data by the program SOLVE and the R_{merge} values, which reflect how good the data merge and also gives an idea whether the heavy atom might be bound to the crystals or not. In general, R_{merge} values below 0.10 could be an indication that the heavy atoms might not be bound. Values between 0.10- 0.20 could be an indication that the heavy atom is bound. R_{merge} values above 0.20 could be due to non-isomorphous nature of the two crystals. These values however are typical values but can vary a little based on the individual experimental conditions and quality of data etc. The R_{merge} values are computed in the table against the resolution range of 9.91Å-4.46Å.

It can be seen that five derivatives tantalum bromide cluster, Potassium hexabromoplatinate (designated Pt1), gold (III)chloride (designated Au), potassium platinum iodide (designated Pt2), and potassium hexabromoplatinate (designated Pt3) gives reasonable figures of merit (a reasonable figure of merit was considered to be 0.20 in this case) over the resolution range of 9.91Å - 4.46Å. The SIR phases calculated by these derivatives were used in map calculation and it was seen that the electron density in the map was not interpretable. Density modification procedures which are built in the program

RESOLVE were carried out to see if the maps can be improved. These maps, even after density modification procedures were not interpretable.

The heavy atom searches were carried out again using the derivatives which gave figures of merit (FOM) higher than 0.20 and were used in MIRAS phasing experiment. The five derivatives gave a combined mean figure of merit of 0.57 over the resolution range of 9.91Å – 4.46Å. Individual figure of merit distribution for various resolution ranges are given in Table 22. Seven heavy atom sites were found by difference Patterson maps by the program SOLVE. The coordinates, occupancies and the B factors are given for each heavy atom are given in Table23.

Table 21. Results of SIR phasing experiment for individual derivatives of AC.

Derivative	Mean FOM	R_{merge} with native set	Resolution range (Å)	Heavy atom sites				
				x	y	Z	occup--ancy	B factor
Tantalum bromide	0.21	0.210	9.91 - 4.46	56.3	69.9	63.5	1.02	60.0
				99.8	96.4	49.8	0.02	10.0
Cluster				10.1	14.0	43.7	0.98	60.0
Potassium Hexabromo platinate (IV)	0.35	0.190	8.81 - 4.55	57.8	36.5	39.3	0.69	51.0
				67.8	112.5	37.3	0.64	60.0
Gold (III) Chloride	0.22	0.182	9.91 - 4.46	51.4	109.7	39.7	0.67	46.6
				51.4	33.7	37.8	0.66	43.8
Potassium Tetrachloroplatinate	0.27	0.180	9.91 - 4.46	57.8	36.5	39.3	0.69	51.0
				57.8	112.5	37.3	0.64	60.0
Mercury (II)chloride	0.10	0.120	9.90 - 4.47	22.0	32.0	12.0	0.09	65.0
Sodium Tungstate dihydrate	0.08	0.140	9.91 - 4.46	13.0	67 .7	11.9	0.03	60.0
Sodium Bromide	0.09	0.130	14.1 - 7.61	18.3	99.9	47.5	1.02	60.0
				14.3	79.9	33.5	0.39	60.0

Table 21. Results of SIR phasing experiment for individual derivatives of AC-continued.

Platinum potassium iodide	0.22	0.160	9.70	- 4.55	51.4 51.4	109.7 33.7	39.7 37.8	0.67 0.66	46.6 43.8

Table 22. Distribution of figures of merit (FOM) versus resolution.

Resolution range (Å)	Figure of merit (FOM)
9.91-6.26	0.57
6.26-5.62	0.52
5.62-4.86	0.50
4.86-4.46	0.48

Table 23. Results of MIR phasing experiment for individual derivatives of AC.

Resolution range (Å)	Phasing power	Heavy atom sites					
			x	y	z	Occupancy	B factor Å^2
9.91-6.26	0.81	Ta	58.9	66.8	65.9	1.27	60.0
6.26-5.62	1.07	Ta	54.1	145.6	65.3	0.58	36.5
5.62-4.86	0.93						
4.86-4.46	0.77						
9.91-6.26	0.83	Pt1a	61.8	67.6	65.4	0.89	60.0
6.26-5.62	0.88	Pt1b	59.2	37.1	42.3	0.69	56.0
5.62-4.86	0.69						
4.86-4.46	0.52						
9.91-6.26	0.92	Au	60.8	67.3	65.3	1.02	60.0
6.26-5.62	1.12						
5.62-4.86	0.77						
4.86-4.46	0.68						
9.91-6.26	0.84	Pt2	61.0	23.8	65.4	0.87	48.6
6.26-5.62	0.92						
5.62-4.86	0.75						
4.86-4.46	0.59						
9.91-6.26	0.9	Pt3	61.3	67.4	65.3	0.85	59.2
6.26-5.62	1.0						
5.62-4.86	0.88						
4.86-4.46	0.67						

The quality of the solution obtained from MIRAS or MAD after heavy atom search can be described using many two important parameters; figures of merit (FOM) and phasing power. Figure of merit is an appropriate weight calculated from the phase probability distribution to weigh down certain reflections which have uncertain phases. Uncertain phases will result in erroneous electron density maps and have to be weighted down with an appropriate weight. The closer the value of FOM to 1, the better is the probability of the determined phase angle and values tending towards zero means that the phases are not accurately determined. The mean figure of merit versus resolution (Table 22) gives an estimation of the accuracy of phase determination at different resolutions.

The other parameter which could be quoted as an evidence to show that a particular derivative is a useful one is the phasing power. Phasing power compares the magnitude of the calculated heavy atom structure factor (F_H) to the lack of closure. When calculating the structure factor of the heavy-atom bound protein, (F_{PH}), there will be errors. The lack of closure is the measure of this error. The lower the error, the more accurate is an assumed phase angle. Values of the phasing power > 1.5 are considered excellent, > 1 good, > 0.5 usable. A plot of phasing power versus resolution gives an idea of how useful a derivative is in phasing at different resolutions.

Using the MIR phasing it can be seen that the overall figure of merit obtained from the 5 derivatives of AC is around 0.55. Although the figures of

merit (FOM) are reasonably higher in lower resolution shell, they considerably drop at around 6 Å and reach 0.48 at 4.5 Å. It can be seen from Table 23 that the derivatives provide only moderate phasing power at lower resolutions and are very poor at higher resolutions. A further problem with these derivatives was that many of the derivatives with the best phasing bind precisely at the same site on the protein and hence do not provide independent information (one of the heavy atom sites designated Pt1a, the heavy atom site designated Au and the site designated Pt2). The phase information generated by the 5 derivatives was poor beyond 4.5 Å and a map generated by MIRAS was not interpretable and further procedures were required before interpretable maps could be obtained.

Density Modification Procedures

Perhaps the most important strategy to improve electron density maps is to obtain a good set of derivatives capable of producing a good structure solution. Before doing that, there are considerable numbers of ways to improve electron density maps using the existing set of data. Such methods are collectively referred to as density-modification (DM). Some of the examples of density modifications are NCS averaging, histogram matching and solvent flattening.

Finding the Non Crystallographic Symmetry

A solvent content of 60% [126] indicates that a complete hexamer is present in the asymmetric unit for the above mentioned crystal forms of AC. Since the hexamer consists of two of each of α, β and γ subunits related by a two

fold internal symmetry, the presence of an entire hexamer in the asymmetric unit allows the use of non crystallographic symmetry to aid in improving of the electron density maps. Use of NCS has been proven to be powerful in improving poorly interpretable electron density maps [127]. The program RESOLVE utilizes the coordinates of the heavy atoms to find NCS. The heavy atom coordinates from the MIR experiment (Table 23) were provided to the program. Initially the program could not find the two fold NCS with these coordinates. Previously we have been able to find NCS for other crystal structures using a heavy atom file obtained by combining heavy atom sites obtained from a number of searches carried out using only one derivative at a time with the heavy atom sites got by searches conducted on these derivatives combined together. By providing such coordinate file to the program RESOLVE [82] a two fold NCS operation was identified. The program carries out the density modification procedure such as solvent flattening and also carries out automated model building where it builds and fits residues of the protein molecule into the calculated density if the sequence of the protein is provided. The automated model building procedure built about 25 residues out of which 200 residues were alanine residues which indicated that the program couldn't build side chains into the density calculated. The density modification and the automated building procedure of the program RESOLVE improved the figure of merit from 0.57 to 0.67 over the resolution range of 9.91 to 4.6Å. The program also calculates new phases after the density modification procedures and model building. The

electron density map obtained after RESOLVE was still not interpretable and needed further modifications to provide an improved map.

Phase Extension, Solvent Fattening and Histogram Modification

Since the overall figure of merit was good up to 4.46Å resolution and NCS was identified, the next step towards the improvement of electron density maps was phase extension. Table 18 indicates that for the gold chloride derivative, data set is available to 2.7 Å. This data set was used for phase extension using the DM module of CCP4 program [144]. The starting phases were those obtained after NCS averaging. The overall figure of merit for these phases was 0.67 at 9.91Å-4.46Å resolution range. Phase extension was carried out for 275 cyclesfrom a staring resolution of 5Å to 2.7Å. This phase extension was combined with other density modification procedures such as solvent flattening [145], NCS averaging [146] and histogram matching [147]. After density modification procedures, the electron density maps were obtained using the newly calculated phases. The final correlation coefficient for the calculated NCS was 0.95. The figure of merit for the newly calculated phases was 0.72 for a resolution range of 50-2.7Å suggesting improvements to the phase calculations. These phases were used in the calculation of a map at 2.7Å resolution and this map clearly showed a protein envelope relative to solvent (Figure 56). The quality of maps was assessed to see if the protein structure can be interpreted. From the maps it was seen the secondary structures of the protein are not

interpretable from these maps. Nevertheless, the discrimination of solvent from the protein is a significant step towards providing a structure solution. Interpretable electron density maps could be obtained by obtaining improved phases at higher resolutions than those available at this point of time.

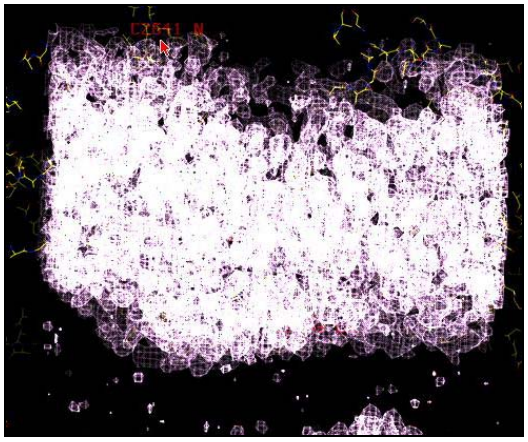

Figure 50. Initial maps contoured at 1σ cutoff revealing a clear solvent-protein envelope.

Data Collection and Experiments on Native and Derivatized Crystals of Acetone Carboxylase belonging to Space Group $P2_1$.

Acetone carboxylase occasionally crystallizes in the space group $P2_1$ as mentioned earlier. Also crystals grown in the presence of mercury salts belong to the same space group. Both single and multiple wavelength data sets were collected on these crystals. The data statistics and details of the calculation of phases from these data sets by MAD, SIR and MIR methods experiments conducted these crystals have been reported in the next section.

MAD Phasing with Acetone Carboxylase Crystals
Grown in Presence of Mercury Salts

The crystals of acetone carboxylase which grew in presence of mercury salts were harvested and flash frozen in liquid nitrogen and subjected to fluorescence scan to obtain energies corresponding to peak, inflection and remote wavelengths and a three wavelength data set was collected. The data statistics are provided in Table 24.

Table 24. Multiple wavelength anomalous diffraction data statistics for crystals Grown in presence of mercury (III) chloride.

Space group	$P2_1$		
Cell dimensions	a = 76.1 Å, b = 152..2 Å c = 264.2 Å, α = β= γ = 91.2°		
	Peak λ=1.0794 Å	Remote λ= 0.89194 Å	Inflection λ=1.07222 Å
Resolution (Å)	50.0-2.60	50.0-2.9	50.0-2.9
Completeness (%)	99.2 (99.9)	98.8 (99.2)	96.9 (99.1)
Observed reflections	524322	197822	14\24062
Unique reflections	74215	28928	24438
I/σ	15 (8.0)	14.1 (10)	12.1 (9.5)
R_{merge}[a]	0.106 (0.125)	0.09 (0.120)	0.097 (0.13)

The data sets were processed using HKL2000 and after scaling the different wavelengths separately it was found that the anomalous signal was not strong (indicated by low differences in chi^2 values between high and low resolution shells). Nevertheless the data sets were subjected to heavy atom search by the program SOLVE. The coordinates, occupancies and B factors of the heavy atom sites are given in Table 25. From the table it can also be seen

that the figures of merit are very low for this MAD experiment. The maps calculated from the phases calculated from this experiment were not interpretable even after density modification procedures such as solvent flattening and histogram matching.

Table 25. Results of MAD phasing experiment for mercury (III) chloride derivative of AC.

Resolution range (Å)	FOM	Heavy atom sites					
			x	Y	z	Occupancy	B factor Å^2
8.81-5.75	0.17	1	21.7	121.2	64.0	0.09	7.4
5.75-4.55	0.16	2	23.7	120.2	63.9	0.06	6.8
4.55- 3.88	0.14	3	23.2	21.1	60.3	0.06	43.5
3.88-3.44	0.13	4	25.4	120.2	63.3	0.09	49.9
3.44-3.12	0.10	5	6.3	120.8	63.6	0.06	39.2
3.12-2.88	0.09						

Analysis of Native Data Sets Collected on
Crystals belonging to Space Group $P2_1$

The crystals were harvested on a rayon loop and flash frozen and complete data sets were collected at high energy (12,000eV). The data statistics for two native crystals are given in Tables 26 and 27. When these data sets were merged it could be seen that R_{merge} are considerably low at all resolution ranges of data available suggesting that these crystals are relatively isomorphous (Table 28). These values of R_{merge} are lower than those for $P2_12_12_1$ crystals. This table suggests that the crystals can be considered isomorphous over a resolution range if 50Å to 3.4Å

Table 26. Data statistics of crystals of AC with no heavy atoms (native set 1).

Wavelength Used (Å)	1.2000
Space Group	$P2_1$
Cell dimensions	a=76.3, b=120.8, c=264.0 $\alpha = \beta = \gamma = 91.3^o$
Resolution (Å)	50-3.3
Number of Reflections (number of unique reflections)	214130 (53147)
*R*sym	0.066 (0.099)
% Completeness	98.5(99.1)

Table 27. Data statistics of crystals of AC with no heavy atoms (native set 2).

Wavelength Used (Å)	1.2000
Space Group	$P2_1$
Cell dimensions	a=75.3 b=121.1 c=264.4 $\alpha = \beta = \gamma = 91.5^o$
Resolution (Å)	50-3.0
Number of Reflections (number of unique reflections)	234120 (43127)
*R*sym	0.078 (0.10)
% Completeness	99.5(99.1)

Table 28. R_{merge} as a function of resolution for two native data sets (1 and 2).

Resolution bins (Å)	R_{merge}
50.00 6.89	0.04
6.89 5.47	0.05
5.47 4.78	0.05
4.78 4.34	0.06
4.34 4.03	0.04
4.03 3.79	0.07
3.79 3.60	0.09
3.60 3.45	0.11
3.45 3.31	0.13
3.31 3.15	0.15
3.15 3.01	0.17

The native crystals $P2_1$crystals were soaked in different heavy atom solutions as explained earlier (listed in Table 9) using the dialysis methods and the crystals capable of diffraction were subjected to data collection at different wavelengths

corresponding to the peak energies of the anomalously scattering heavy atom present used in the soaking experiment. The data statistics for the heavy atom soaked crystals are shown in Table 29. The crystals soaked in mercury (III) chloride (designated Hg1), platinum potassium iodide (designated Pt1), potassium hexabromoosmate and EMTS(designated Hg2) were capable of diffraction even after soaking and were used for data collection.

Table 29. Data statistics of crystals of AC soaked in various heavy atoms.

Heavy atom	Mercury (III) chloride	Platinum Potassium iodide	Potassium hexabromoosmate	EMTS
Wavelength Used (Å)	0.99674	1.0587	1.0480	0.99480
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Resolution (Å)	50-2.7	50-2.9	50-3.3	50-3.2
Cell Dimensions (Å)	a=76.2 b=120.8 c=264.1 β=91.3°	a=76.2 b=121.3 c=264.2 β= 91.4°	a=75.2 b=120.2 c=265.1 β=91.2°	a=74.2 b=121.3 c=265.2 β= 91.5°
No. of reflections (unique reflections)	382851 (121412)	424651 (107822)	235442 (43821)	255230 (42848)
*R*sym	0.05 (0.12)	0.08 (0.24)	0.07 (0.13)	0.07 (0.14)
% Completeness	99.5 (97.7)	99.4 (98.3)	99.2 (97.9)	98.1 (97.3)

Heavy atom searches were conducted by utilizing the program SOLVE and the coordinates, occupancies and B factors for the heavy atoms are provided in Table 30. Also, the over all FOM and the corresponding resolution ranges are provided in the table. It can be seen that these derivatives when used in SIR phase calculations do not provide good phase information as reflected by the low

FOM values. Also, the maps calculated using the SIR phases were not interpretable.

Heavy atom searches using all these derivatives were carried out and MIR phases were calculated. The heavy atom coordinates, their occupancies and B factors are shown in Table 31. It can be seen from the same table that these heavy atoms do not provide good phasing power and the mean FOM value of 0.22 over the resolution range of 9.91Å – 2.8Å reflects on phase calculation which are not accurate. Maps calculated utilizing these phases were not interpretable even after solvent flattening and histogram matching.

Table 30. Results of SIR phasing experiment for individual derivatives of AC.

Derivative	Mean FOM	R_{merge} with native set	Resolution range (Å)	Heavy atom sites				
				X	Y	Z	Occup-ancy	B factor Å^2
Mercury (III) chloride	0.19	0.150	9.91 - 2.96	26.3 79.8	39.9 26.4	13.5 149.3	0.12 0.11	60.0 10.0
Platinum Potassium iodide	0.16	0.110	8.81 -3.00	27.8 167.8	136.5 12.5	49.1 77.2	0.19 0.14	55.0 60.0
Potassium hexabromoosmate	0.12	0.122	9.91 - 3.5	51.4 22.4	109.7 33.7	139.7 137.8	0.17 0.16	55.6 45.8
EMTS	0.27	0.090	9.91 - 3.5	27.2 57.8	138.5 112.5	39.3 17.3	0.29 0.24	60.0 60.0

Table 31. Results of MIR phasing experiment for individual derivatives of AC.

Resolution range (Å)	Phasing power	Heavy atom sites					
			X	Y	Z	Occupancy	B factor Å^2
8.81-5.75	0.71	Hg1	68.9	66.8	65.9	1.02	60.0
5.75-4.55	0.57	Hg1	24.1	35.6	15.3	0.58	56.5
4.55-3.88	0.43						
3.88-3.44	0.37						
3.44-3.12	0.35						
3.12-2.88	0.30						
8.81-5.75	0.63	Pt1	51.8	63.6	55.4	0.79	60.0
5.75-4.55	0.58	Pt1	69.2	27.3	32.2	0.69	56.0
4.55-3.88	0.31						
3.88-3.44	0.22						
3.44-3.12	0.11						
3.12-2.88	0.10						
8.81-5.75	0.92	Hg2	60.8	67.3	65.3	0.82	60.0
5.75-4.55	0.82	Hg2	62.3	25.2	23.8	0.91	60.0
4.55-3.88	0.77						
3.88-3.44	0.62						
3.44-3.12	0.55						
3.12-2.88	0.23						

Use of ATP Analogues as Heavy Atom Derivatives

It is already known that AC requires ATP for catalysis. This is a favorable condition for generating an ATP based heavy atom derivative of the protein. 2'-deoxy-2'-iodoadenosine-5'-triphosphate (2'-IATP) has been successfully used to derivatize many protein crystals for MAD and MIR experiments [142]. As the name indicates 2'-IATP has an iodine atom at the 2' position of the ribose sugar. Since the enzymes generally bind with equal affinities to 2'-IATP and ATP, and there is no chemistry happening at this site, it is a convenient tool to incorporate heavy atoms into the crystals. Halides like iodide have proven to be very useful in the determination of structures of many

proteins [143]. A 5 mM solution of 2'-IATP was made with the mother liquor and the crystals were soaked in this solution for different amounts of time varying from 5 minutes to 12 hours. The crystals soaked for up to 8 hours did not lose the ability to diffract and were selected for data collection experiments. The crystals were harvested using rayon loops and were then flash frozen under liquid nitrogen and utilized for data collection. A single wavelength data set was collected at the absorption edge of iodine. After data processing it was seen that this crystal belonged to the space group $P2_1$. The data statistics for a 2-IATP soaked crystal are given in Table 32. The data set of 2'IATP was scaled with the native data set (set 1) shown in Table 26 and the scaling statistics are provided in table. It can be seen from the R_{merge} value from the two data sets that the possibility of ATP being bound is not high as reflected by the low R_{merge} values (Table 33).

Table 32. Data statistics of crystals of AC soaked in 2'-Iodo ATP.

Wavelength Used (Å)	1.2000
Space Group	$P2_1$
Cell dimensions	a=76.3 b=120.1 c=264.7 β= 91.2°
Resolution (Å)	50-3.0
Number of Reflections (number of unique reflections)	204020 (41115)
*R*sym	0.098 (0.15)
% Completeness	96.5(97.1)

Table 33. R_{merge} as a function of resolution.

Resolution bins (Å)	R_{merge}
50.0 - 6.89	0.03
6.89 - 5.47	0.05
5.47 - 4.78	0.05
4.78 - 4.34	0.06
4.34 - 4.03	0.04
4.03 - 3.79	0.07
3.79 - 3.60	0.08
3.60 - 3.45	0.09
3.45 - 3.31	0.11
3.31 - 3.15	0.12
3.15 - 3.01	0.12

If initial phase information for the crystal for $P2_1$ was available, difference Fourier maps could have been calculated to confirm the presence or absence of 2'-IATP in the crystals soaked with 2'-IATP. Unfortunately at this time this is not possible because there are no initial phases available for the $P2_1$ crystal form.

Summary, Conclusions and Future Directions

We have realized that there are several impediments to solving the structure of acetone carboxylase. With no homology to other enzymes, molecular replacement is not an option at present. Since the protein cannot be heterologously expressed, it is not possible to obtain a selenomethionine derivative. Further, the non isomorphous nature of the crystals has made MIR difficult. Because of the significant size of the enzyme, there is a possibility of incorporating a large number of heavy atoms which poses problems during the heavy atom search. Even with all the aforementioned impediments there has been a significant amount of advancement towards obtaining a structure solution

of AC. These studies have also resulted in optimizing conditions for obtaining reasonably good derivatives. Because of these efforts, good quality data sets for MIR experiments can now be routinely collected. Most importantly, the studies carried out above have resulted in maps where a clear protein-solvent boundary is discernable. Even if the electron density maps are not fully interpretable at this stage, with NCS operator in hand and significant phasing from the existing derivatives, identification of additional derivatives will be much easier by providing the initial phase information. Lastly, this project has great educational value and has provided hands-on experience to learn the fundamental techniques involved in obtaining the structure of proteins using the technique of X-ray crystallography.

Future directions of this project involve obtaining good quality derivatives with isomorphous cell parameters to obtain a structure solution. The maps calculated using the derivatives of acetone carboxylase in the $P2_12_12_1$ crystal form have revealed a solvent envelope. Upon availability of additional derivatives which are better than 4Å resolution in the same space group, the initial phase information that we already have can be utilized and combined with the new phase information got by these new derivatives which could result in better maps where secondary structures are interpretable. Also, the $P2_1$ crystals are relatively non isomorphous. If more derivatives that can help in providing additional phase information can be obtained, new maps can be calculated with this phase information. Once initial electron density maps are available for this

crystal form where a clear solvent boundary can be seen, they can be used in conjunction with the maps available for $P2_12_12_1$ form to get a better quality map. This could be done by carrying out cross crystal averaging between the two crystal forms. There are several programs (Ex. DMULTI, CRAVE) available which can calculate a correlation map between the densities of a molecule as found in two different crystal forms. Cross crystal averaging combined with NCS averaging in these two crystal forms can lead to structure solution.

Once the structure is solved the available data for acetone carboxylase crystals soaked in 2'-IATP could be tested by difference Fourier to confirm presence or absence of 2'-IATP and if it is bound it could provide information about the role of ATP. Once the native structure is solved, substrates, products, substrate analogues, ATP analogues etc can be soaked into the crystals or co-crystallized and by analyzing the interactions between such bound compounds and the enzyme, the mechanism of the enzyme can be understood.

CHAPTER 6

CONCLUDING REMARKS

The overall goal of the dissertation presented above was to provide structural insights into the reaction mechanisms of the enzymes which are involved in the propylene and isopropanol metabolism in *Xanthobacter autotrophicus* which in turn would provide insights in general about alkene and alcohol metabolism by microorganisms in general. The in-depth understanding of the metabolism of alkenes and alcohols will not only provide information about the novel transformations involved in metabolizing these xenobiotics but also enhance the knowledge of the unique carboxylases involved in fixing carbon dioxide in these organisms. Structural analysis of various enzymes involved in the pathway of propylene metabolism is carried out in the work reported here and has made major contributions to the overall understanding of the metabolism of alkenes and epoxides.

The three dimensional structure of *R*-HPCDH has been determined to a resolution of 1.8Å and has been compared with the homology model of S-HPCDH. These results provide a structural basis for chiral discrimination of the *R* and *S*-HPC substrates by the stereo selective dehydrogenase pair and support our previously proposed model. Although the overall structure and architecture of both the enzymes are analogous, there are key differences in the substrate binding sites. Differential placement of positively charged residues that bind the

negatively charged sulfonate group of the substrate in the substrate binding region supports the proper orientation of the hydroxyl and the hydrogen group of the enzyme for proton and hydride abstraction. The sulfonate group is a crucial determinant for orienting the reactive portion of the substrates for chiral discrimination.

The understanding of chiral discrimination and catalysis in the dehydrogenase pair of enzymes has been further advanced by the structural studies conducted on the methionine mutants. The structural studies of the Met192Ala and the Met187Ala have revealed the importance of these methionines in substrate binding and catalysis. They are important in forming a compact solvent inaccessible substrate pocket, in shielding the electrostatic interactions between the substrate and the enzyme and in shielding the charges of the two ends of the substrate molecules from each other.

In the following section of the dissertation, the three dimensional structures where available and the homology models of several enzymes involved in alkene metabolism from various genera of microorganisms like *Xanthobacter, Rhodococcus, Nocardiodes , Pseudomonas , Ochrobactrum* and *Mycobacterium* have been analyzed. These microorganisms are capable of using alkenes and chlorinated alkenes like propylene, ethylene and vinyl chloride as sole sources of carbon. The enzymes involved in metabolism of the aforementioned compounds utilize CoM as the cofactor in the reactions involved in the metabolism. Primary sequence analysis along with multiple sequence alignment has suggested that

these enzymes have similarities in the CoM binding sites involving positively charged residues (Arg/Lys) which are conserved among all the enzymes which bind the sulfonate moiety of CoM. Three dimensional structures and homology models of several enzymes involved in the pathway have been determined and substrate molecules are modeled in their active sites. These structures reveal the binding mode of the sulfonate moiety of the substrates providing clues to the key interactions involved between the enzymes and the substrate. The similarities in the binding sites are striking, suggesting a general signature for coenzymeM dependent enzymes. It emphasizes the fact that the sulfonate group is a crucial determinant for proper binding and orientation of the substrate in the substrate binding site of the enzyme for catalysis and highlights the efficacy of this small cofactor as a carrier molecule in these metabolic pathways.

Finally, a significant amount of progress has been made toward the determination of the structure of acetone carboxylase. Good diffraction quality crystals can now be routinely grown and derivatized for the purpose of MAD and MIR. The current set of data that have been collected on various heavy atom derivatives of AC have allowed us to identify several heavy atom sites and obtain an initial map in which a clear protein-solvent boundary can be discerned. The current set of data will provide groundwork for the future trials of structure determination.

This project has utilized a multidisciplinary approach involving kinetic studies, site specific amino acid substitution studies, homology modeling and

determination of high resolution structures of enzymes in the presence of substrates to ascertain information concerning the biochemical mechanisms of enzyme catalyzed reactions. The results obtained here provide strong basis for the previously conducted kinetic, mutagenic and biochemical studies and form a good compliment to them.

REFERENCES CITED

1. Keppler, F., et al., *Formation of volatile iodinated alkanes in soil: results from laboratory studies.* Chemosphere, 2003. **52**(2): p. 477-83.

2. Kielhorn, J., et al., *Vinyl chloride: still a cause for concern.* Environ Health Perspect, 2000. **108**(7): p. 579-88.

3. Keppler, F., et al., *Natural formation of vinyl chloride in the terrestrial environment.* Environ Sci Technol, 2002. **36**(11): p. 2479-83.

4. Maymo-Gatell, X., et al., *Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene.* Science, 1997. **276**(5318): p. 1568-71.

5. Wade, D.R., S.C. Airy, and J.E. Sinsheimer, *Mutagenicity of aliphatic epoxides.* Mutat Res, 1978. **58**(2-3): p. 217-23.

6. Guengerich, F.P., *Reactions and significance of cytochrome P-450 enzymes.* J Biol Chem, 1991. **266**(16): p. 10019-22.

7. Hartmans, S., J.A. de Bont, and W. Harder, *Microbial metabolism of short-chain unsaturated hydrocarbons.* FEMS Microbiol Rev, 1989. **5**(3): p. 235-64.

8. Armstrong, R.N., *Structure, catalytic mechanism, and evolution of the glutathione transferases.* Chem Res Toxicol, 1997. **10**(1): p. 2-18.

9. Thomas H, T.C., Oesch F, *Principles, Mechanisms and Biological Consequences of Induction*, ed. H.R. K Ruckpaul. 1990: New York: Taylor & Francis. 280-337.

10. Ensign, S.A., *Microbial metabolism of aliphatic alkenes.* Biochemistry, 2001. **40**(20): p. 5845-53.

11. van Ginkel, C., G. de Bont, J,A., Arch Microbiol, 1986. **145**: p. 403-407.

12. van Ginkel, C.G., H.G. Welten, and J.A. de Bont, *Oxidation of Gaseous and Volatile Hydrocarbons by Selected Alkene-Utilizing Bacteria.* Appl Environ Microbiol, 1987. **53**(12): p. 2903-2907.

13. Ensign, S.A., M.R. Hyman, and D.J. Arp, *Cometabolic degradation of chlorinated alkenes by alkene monooxygenase in a propylene-grown Xanthobacter strain.* Appl Environ Microbiol, 1992. **58**(9): p. 3038-46.

14. Ensign, S.A., et al., *New roles for CO2 in the microbial metabolism of aliphatic epoxides and ketones.* Arch Microbiol, 1998. **169**(3): p. 179-87.

15. Knowles, J.R., *The mechanism of biotin-dependent enzymes.* Annu Rev Biochem, 1989. **58**: p. 195-221.

16. Cleland, W.W., et al., *Mechanism of Rubisco: The Carbamate as General Base.* Chem Rev, 1998. **98**(2): p. 549-562.

17. Janc, J.W., M.H. O'Leary, and W.W. Cleland, *A kinetic investigation of phosphoenolpyruvate carboxylase from Zea mays.* Biochemistry, 1992. **31**(28): p. 6421-6.

18. Dowd, P., et al., *Vitamin K and energy transduction: a base strength amplification mechanism.* Science, 1995. **269**(5231): p. 1684-91.

19. Small, F.J. and S.A. Ensign, *Carbon dioxide fixation in the metabolism of propylene and propylene oxide by Xanthobacter strain Py2.* J Bacteriol, 1995. **177**(21): p. 6170-5.

20. Small, F.J., J.K. Tilley, and S.A. Ensign, *Characterization of a New Pathway for Epichlorohydrin Degradation by Whole Cells of Xanthobacter Strain Py2.* Appl Environ Microbiol, 1995. **61**(4): p. 1507-1513.

21. Muira, A., Dalton, H., Biosci Biotechnol Biochem, 1995. **59**: p. 853-859.

22. Small, F.J. and S.A. Ensign, *Alkene monooxygenase from Xanthobacter strain Py2. Purification and characterization of a four-component system*

central to the bacterial metabolism of aliphatic alkenes. J Biol Chem, 1997. **272**(40): p. 24913-20.

23. Gallagher, S.C., A. George, and H. Dalton, *Sequence-alignment modelling and molecular docking studies of the epoxygenase component of alkene monooxygenase from Nocardia corallina B-276.* Eur J Biochem, 1998. **254**(3): p. 480-9.

24. Gallagher, S.C., R. Cammack, and H. Dalton, *Electron transfer reactions in the alkene mono-oxygenase complex from Nocardia corallina B-276.* Biochem J, 1999. **339 (Pt 1)**: p. 79-85.

25. Gallagher, S.C., R. Cammack, and H. Dalton, *Alkene monooxygenase from Nocardia corallina B-276 is a member of the class of dinuclear iron proteins capable of stereospecific epoxygenation reactions.* Eur J Biochem, 1997. **247**(2): p. 635-41.

26. Elango, N., et al., *Crystal structure of the hydroxylase component of methane monooxygenase from Methylosinus trichosporium OB3b.* Protein Sci, 1997. **6**(3): p. 556-68.

27. Rosenzweig, A.C., et al., *Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane.* Nature, 1993. **366**(6455): p. 537-43.

28. Zhou, N.Y., et al., *The alkene monooxygenase from Xanthobacter strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol.* Appl Environ Microbiol, 1999. **65**(4): p. 1589-95.

29. Allen, J.R. and S.A. Ensign, *Two short-chain dehydrogenases confer stereoselectivity for enantiomers of epoxypropane in the multiprotein epoxide carboxylating systems of Xanthobacter strain Py2 and Nocardia corallina B276.* Biochemistry, 1999. **38**(1): p. 247-56.

30. Allen, J.R., et al., *A role for coenzyme M (2-mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8432-7.

31. Allen, J.R. and S.A. Ensign, *Carboxylation of epoxides to beta-keto acids in cell extracts of Xanthobacter strain Py2.* J Bacteriol, 1996. **178**(5): p. 1469-72.

32. Allen, J.R. and S.A. Ensign, *Purification to homogeneity and reconstitution of the individual components of the epoxide carboxylase multiprotein enzyme complex from Xanthobacter strain Py2.* J Biol Chem, 1997. **272**(51): p. 32121-8.

33. Allen, J.R. and S.A. Ensign, *Characterization of three protein components required for functional reconstitution of the epoxide carboxylase multienzyme complex from Xanthobacter strain Py2.* J Bacteriol, 1997. **179**(10): p. 3110-5.

34. Allen, J.R. and S.A. Ensign, *Identification and characterization of epoxide carboxylase activity in cell extracts of Nocardia corallina B276.* J Bacteriol, 1998. **180**(8): p. 2072-8.

35. Wolfe, R.S., *My kind of biology.* Annu Rev Microbiol, 1991. **45**: p. 1-35.

36. Taylor, C.D., et al., *Coenzyme M, essential for growth of a rumen strain of Methanobacterium ruminantium.* J Bacteriol, 1974. **120**(2): p. 974-5.

37. Thauer, R.K., *Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture.* Microbiology, 1998. **144 (Pt 9)**: p. 2377-406.

38. Clark, D.D., J.R. Allen, and S.A. Ensign, *Characterization of five catalytic activities associated with the NADPH:2-ketopropyl-coenzyme M [2-(2-ketopropylthio)ethanesulfonate] oxidoreductase/carboxylase of the Xanthobacter strain Py2 epoxide carboxylase system.* Biochemistry, 2000. **39**(6): p. 1294-304.

39. Swaving, J., et al., *Complementation of Xanthobacter Py2 mutants defective in epoxyalkane degradation, and expression and nucleotide sequence of the complementing DNA fragment.* Microbiology, 1995. **141 (Pt 2)**: p. 477-84.

40. Krum, J.G. and S.A. Ensign, *Heterologous expression of bacterial Epoxyalkane:Coenzyme M transferase and inducible coenzyme M biosynthesis in Xanthobacter strain Py2 and Rhodococcus rhodochrous B276.* J Bacteriol, 2000. **182**(9): p. 2629-34.

41. Saderholm, M.J., K.E. Hightower, and C.A. Fierke, *Role of metals in the reaction catalyzed by protein farnesyltransferase.* Biochemistry, 2000. **39**(40): p. 12398-405.

42. Huang, C.C., P.J. Casey, and C.A. Fierke, *Evidence for a catalytic role of zinc in protein farnesyltransferase. Spectroscopy of Co2+-farnesyltransferase indicates metal coordination of the substrate thiolate.* J Biol Chem, 1997. **272**(1): p. 20-3.

43. Myers, L.C., et al., *Repair of DNA methylphosphotriesters through a metalloactivated cysteine nucleophile.* Science, 1993. **261**(5125): p. 1164-7.

44. Myers, L.C., G.L. Verdine, and G. Wagner, *Solution structure of the DNA methyl phosphotriester repair domain of Escherichia coli Ada.* Biochemistry, 1993. **32**(51): p. 14089-94.

45. Breksa, A.P., 3rd and T.A. Garrow, *Recombinant human liver betaine-homocysteine S-methyltransferase: identification of three cysteine residues critical for zinc binding.* Biochemistry, 1999. **38**(42): p. 13991-8.

46. Thanbichler, M., B. Neuhierl, and A. Bock, *S-methylmethionine metabolism in Escherichia coli.* J Bacteriol, 1999. **181**(2): p. 662-5.

47. Sauer, K. and R.K. Thauer, *Methyl-coenzyme M formation in methanogenic archaea. Involvement of zinc in coenzyme M activation.* Eur J Biochem, 2000. **267**(9): p. 2498-504.

48. Boyd, J.M. and S.A. Ensign, *Evidence for a metal-thiolate intermediate in alkyl group transfer from epoxypropane to coenzyme M and cooperative metal ion binding in epoxyalkane:CoM transferase.* Biochemistry, 2005. **44**(39): p. 13151-62.

49. Peariso K., C.D., Krum JG., ensign SA., Penner-Hahn JE, j biol inorg chem, 2001. **manuscript submitted**.

50. Boyd, J.M., H. Ellsworth, and S.A. Ensign, *Bacterial acetone carboxylase is a manganese-dependent metalloenzyme.* J Biol Chem, 2004. **279**(45): p. 46644-51.

51. Clark, D.D. and S.A. Ensign, *Characterization of the 2-[(R)-2-hydroxypropylthio]ethanesulfonate dehydrogenase from Xanthobacter strain Py2: product inhibition, pH dependence of kinetic parameters, site-directed mutagenesis, rapid equilibrium inhibition, and chemical modification.* Biochemistry, 2002. **41**(8): p. 2727-40.

52. Jornvall, H., et al., *Short-chain dehydrogenases/reductases (SDR).* Biochemistry, 1995. **34**(18): p. 6003-13.

53. Persson, B., M. Krook, and H. Jornvall, *Characteristics of short-chain alcohol dehydrogenases and related enzymes.* Eur J Biochem, 1991. **200**(2): p. 537-43.

54. Wierenga, R.K., P. Terpstra, and W.G. Hol, *Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint.* J Mol Biol, 1986. **187**(1): p. 101-7.

55. Kallberg, Y., et al., *Short-chain dehydrogenases/reductases (SDRs).* Eur J Biochem, 2002. **269**(18): p. 4409-17.

56. Rossmann, M., Liljas, A branden, C banaszak, LJ, in *The enzymes, Academic Press, Newyork*. 1975. p. 61-102.

57. Tanaka N, N.T., Nakamura K.T, Hara A.Current, *SDR Structure, Mechanism of Action, and Substrate Recognition.* Organic Chemistry, (2001): p. pp. 89-111.

58. Yamamoto, K., et al., *Crystal structure of glucose dehydrogenase from Bacillus megaterium IWG3 at 1.7 A resolution.* J Biochem (Tokyo), 2001. **129**(2): p. 303-12.

59. Kleywegt, G.J. and T.A. Jones, *Template convolution to enhance or detect structural features in macromolecular electron-density maps.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 2): p. 179-85.

60. Horer, S., et al., *The crystallographic structure of the mannitol 2-dehydrogenase NADP+ binary complex from Agaricus bisporus.* J Biol Chem, 2001. **276**(29): p. 27555-61.

61. Nakajima, K., T. Hashimoto, and Y. Yamada, *Two tropinone reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor.* Proc Natl Acad Sci U S A, 1993. **90**(20): p. 9591-5.

62. Nakajima, K., T. Hashimoto, and Y. Yamada, *Opposite stereospecificity of two tropinone reductases is conferred by the substrate-binding sites.* J Biol Chem, 1994. **269**(16): p. 11695-8.

63. Nakajima, K., et al., *Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 4876-81.

64. Nakajima, K., et al., *Site-directed mutagenesis of putative substrate-binding residues reveals a mechanism controlling the different stereospecificities of two tropinone reductases.* J Biol Chem, 1999. **274**(23): p. 16563-8.

65. Yamashita, A., et al., *Structure of tropinone reductase-II complexed with NADP+ and pseudotropine at 1.9 A resolution: implication for stereospecific substrate binding and catalysis.* Biochemistry, 1999. **38**(24): p. 7630-7.

66. Nocek, B., et al., *Structural basis for CO2 fixation by a novel member of the disulfide oxidoreductase family of enzymes, 2-ketopropyl-coenzyme M oxidoreductase/carboxylase.* Biochemistry, 2002. **41**(43): p. 12907-13.

67. Clark, D.D., J.M. Boyd, and S.A. Ensign, *The stereoselectivity and catalytic properties of Xanthobacter autotrophicus 2-[(R)-2-Hydroxypropylthio]ethanesulfonate dehydrogenase are controlled by*

interactions between C-terminal arginine residues and the sulfonate of coenzyme M. Biochemistry, 2004. **43**(21): p. 6763-71.

68. Pai, E.F., Curr Opin Struct Biol, 1991(1): p. 796-803.

69. Karplus, P.A., E.F. Pai, and G.E. Schulz, *A crystallographic study of the glutathione binding site of glutathione reductase at 0.3-nm resolution.* Eur J Biochem, 1989. **178**(3): p. 693-703.

70. Kuriyan, J., et al., *X-ray structure of trypanothione reductase from Crithidia fasciculata at 2.4-A resolution.* Proc Natl Acad Sci U S A, 1991. **88**(19): p. 8764-8.

71. Schierbeek, A.J., et al., *X-ray structure of lipoamide dehydrogenase from Azotobacter vinelandii determined by a combination of molecular and isomorphous replacement techniques.* J Mol Biol, 1989. **206**(2): p. 365-79.

72. Argiles, J.P., *Has acetone a role in he conversion of fat to carbohydrate in mammals?* Trends Biochem Sci., 1986. **11**: p. 61-65.

73. Kalapos, M.P., *On the mammalian acetone metabolism: from chemistry to clinical implications.* Biochim Biophys Acta, 2003. **1621**(2): p. 122-39.

74. Sluis, M.K. and S.A. Ensign, *Purification and characterization of acetone carboxylase from Xanthobacter strain Py2.* Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8456-61.

75. Sluis, M.K., et al., *Involvement of an ATP-dependent carboxylase in a CO2-dependent pathway of acetone metabolism by Xanthobacter strain Py2.* J Bacteriol, 1996. **178**(14): p. 4020-6.

76. Nordling, E., H. Jornvall, and B. Persson, *Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling.* Eur J Biochem, 2002. **269**(17): p. 4267-76.

77. Sluis, M.K., et al., *Biochemical, molecular, and genetic analyses of the acetone carboxylases from Xanthobacter autotrophicus strain Py2 and Rhodobacter capsulatus strain B10.* J Bacteriol, 2002. **184**(11): p. 2969-77.

78. Archelas, A. and R. Furstoss, *Synthesis of enantiopure epoxides through biocatalytic approaches.* Annu Rev Microbiol, 1997. **51**: p. 491-525.

79. Nocek, B., et al., *Crystallization and preliminary X-ray analysis of an R-2-hydroxypropyl-coenzyme M dehydrogenase.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 9): p. 1470-3.

80. Garman, E.F., Mitchell, E.P, J. Appl. Cryst., 1996. **29**: p. 584- 587.

81. Otwinowski, Z., and Minor, W., *in Methods in Enzymology (Carter, C. W., Jr., and Sweet, R. M., Eds.) p 20, Academic Press,New York.* 1997.

82. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D Biol Crystallogr, 1999. **55 (Pt 4)**: p. 849-61.

83. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.

84. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54 (Pt 5)**: p. 905-21.

85. Brunger, A.T., *The free R value: a novel statistical quantity for assessing the accuracy of crystal structures.* nature, 1992. **355**: p. 472-474.

86. Laskowski R A, M.M.W., Moss D S & Thornton J M J. Appl. Cryst., 26, 283-291., *PROCHECK: a program to check the stereochemical quality of protein structures.* (1993).

87. Sogabe, S., et al., *The crystal structure and stereospecificity of levodione reductase from Corynebacterium aquaticum M-13.* J Biol Chem, 2003. **278**(21): p. 19387-95.

88. Lund, O., Nielsen, M., Lundegaard, C., Worning, P., *X3M a Computer Program to Extract 3D Models.* Abstract at the CASP5 conference A102, 2002.

89. Gasteiger, E., et al., *ExPASy: The proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Res, 2003. **31**(13): p. 3784-8.

90. Guex, N., and Peitsch, M. C., *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. **18**: p. 2714-23.

91. http://www.povray.org/povlegal.html.

92. Rossmann, M.G., Liljas, A., Branden, C.-I., and Banaszak, L. J., *in evolutionary and structural relationships among dehydrogenases (Boyer, P.D., ed) Academic Press, New York.* 1975.

93. Tanaka N, N.T., Nakamura K.T, Hara A., *SDR Structure, Mechanism of Action, and Substrate Recognition.* Current Organic Chemistry, 2001. **pp. 89-111(23)**.

94. de Jong, R.M., et al., *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, 2003. **22**(19): p. 4933-44.

95. Bottoms, C.A., P.E. Smith, and J.J. Tanner, *A structurally conserved water molecule in Rossmann dinucleotide-binding domains.* Protein Sci, 2002. **11**(9): p. 2125-37.

96. Krishnakumar, A.M., et al., *Structural basis for stereoselectivity in the (R)- and (S)-hydroxypropylthioethanesulfonate dehydrogenases.* Biochemistry, 2006. **45**(29): p. 8831-40.

97. McBride, B.C. and R.S. Wolfe, *A new coenzyme of methyl transfer, coenzyme M.* Biochemistry, 1971. **10**(12): p. 2317-24.

98. Graham, D.E. and R.H. White, *Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics.* Nat Prod Rep, 2002. **19**(2): p. 133-47.

99. Graham, D.E., H. Xu, and R.H. White, *Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate-biosynthesizing enzymes.* J Biol Chem, 2002. **277**(16): p. 13421-9.

100. Matthews, R.G. and C.W. Goulding, *Enzyme-catalyzed methyl transfers to thiols: the role of zinc.* Curr Opin Chem Biol, 1997. **1**(3): p. 332-9.

101. Coleman, N.V. and J.C. Spain, *Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading Mycobacterium strains.* Appl Environ Microbiol, 2003. **69**(10): p. 6041-6.

102. Mattes, T.E., et al., *Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in Nocardioides sp. strain JS614.* Arch Microbiol, 2005. **183**(2): p. 95-106.

103. Danko, A.S., et al., *Involvement of coenzyme M during aerobic biodegradation of vinyl chloride and ethene by Pseudomonas putida strain AJ and Ochrobactrum sp. strain TD.* Appl Environ Microbiol, 2006. **72**(5): p. 3756-8.

104. Coleman, N.V. and J.C. Spain, *Epoxyalkane: coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of mycobacterium strain JS60.* J Bacteriol, 2003. **185**(18): p. 5536-45.

105. Hightower, K.E. and C.A. Fierke, *Zinc-catalyzed sulfur alkyation:insights from protein farnesyltransferase.* Curr Opin Chem Biol, 1999. **3**(2): p. 176-81.

106. Thompson JD, H.D., Gibson TJ., *CLUSTAL W: improving the sensitivity ofprogressive multiple sequence alignment through sequence weighting, position-specific gap penaltiesand weight matrix choice.* Nucleic Acids Res, 1994. **22**(4673-80).

107. Felsenstein, J., *PHYLIP (phylogeny inference package), version 3.5c.* Department of Genetics, University of Washington, Seattle., 1993.

108. Peariso, K., et al., *Characterization of the zinc sites in cobalamin-independent and cobalamin-dependent methionine synthase using zinc and selenium X-ray absorption spectroscopy.* Biochemistry, 2001. **40**(4): p. 987-93.

109. Gonzalez, J.C., et al., *Cobalamin-independent methionine synthase from Escherichia coli: a zinc metalloenzyme.* Biochemistry, 1996. **35**(38): p. 12228-34.

110. Pejchal, R. and M.L. Ludwig, *Cobalamin-independent methionine synthase (MetE): a face-to-face double barrel that evolved by gene duplication.* PLoS Biol, 2005. **3**(2): p. e31.

111. Swaving, J., et al., *A novel type of pyridine nucleotide-disulfide oxidoreductase is essential for NAD+- and NADPH-dependent degradation of epoxyalkanes by Xanthobacter strain Py2.* J Bacteriol, 1996. **178**(22): p. 6644-6.

112. Coleman, J.P. and J.J. Perry, *Purification and characterization of the secondary alcohol dehydrogenase from propane-utilizing Mycobacterium vaccae strain JOB-5.* J Gen Microbiol, 1985. **131**(11): p. 2901-7.

113. Lukins, H.B. and J.W. Foster, *Methyl Ketone Metabolism in Hydrocarbon-Utilizing Mycobacteria.* J Bacteriol, 1963. **85**: p. 1074-87.

114. Janssen, P.H. and B. Schink, *14CO2 exchange with acetoacetate catalyzed by dialyzed cell-free extracts of the bacterial strain BunN grown with acetone and nitrate.* Eur J Biochem, 1995. **228**(3): p. 677-82.

115. Janssen, P.H. and B. Schink, *Metabolic pathways and energetics of the acetone-oxidizing, sulfate-reducing bacterium, Desulfobacterium cetonicum.* Arch Microbiol, 1995. **163**(3): p. 188-94.

116. Cleland WW, A.T., Gutteridge S, Hartman FC, Lorimer GH., *Mechanism of rubisco: The carbamateas general base.* Chem. Rev., 1998(98): p. 549-62.

117. Narindrasorasak S, B.W., *Phosphoenolypyruvate synthetase of Escherichia coli: molecular weight,subunit composition, and*

identification of phosphohistidine in phosphoenzyme intermediate. J Biol Chem, 1977. **252**: p. 3121-7.

118. Hutchins AM, H.J., Adams MW, *Phosphoenolpyruvate synthetase from the hyperthermophilic archaeon Pyrococcus furiosus.* J Bacteriol, 2001. **183**(709-15).

119. Cooper, R.A. and H.L. Kornberg, *Net formation of phosphoenolpyruvate from pyruvate by Escherichia coli.* Biochim Biophys Acta, 1965. **104**(2): p. 618-20.

120. Cooper RA, K.H., *Phophonel pyruvate synthetase.* Meth. Enzymol, 1969. **13**(309-14).

121. Boyd, J.M. and S.A. Ensign, *ATP-dependent enolization of acetone by acetone carboxylase from Rhodobacter capsulatus.* Biochemistry, 2005. **44**(23): p. 8543-53.

122. Cooper, R.A. and H.L. Kornberg, *The mechanism of the phosphoenolpyruvate synthase reaction.* Biochim Biophys Acta, 1967. **141**(1): p. 211-3.

123. Cooper, R.A. and H.L. Kornberg, *The direct synthesis of phosphoenolpyruvate from pyruvate by Escherichia coli.* Proc R Soc Lond B Biol Sci, 1967. **168**(12): p. 263-80.

124. Nocek, B., et al., *Crystallization and preliminary X-ray analysis of an acetone carboxylase from Xanthobacter autotrophicus strain Py2.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 2): p. 385-7.

125. Garman, E.F., Mitchell, E.P, J. Appl. Cryst., 1996. **29**: p. 584- 587.

126. Matthews, B.W., J. Mol. Biol., 1968. **33**: p. 491-497.

127. Blow, D.M. and M.G. Rossmann, *The detection of subunits within the crystallographic assymmetric unit.* Acta Crystallogr D Biol Crystallogr, 1962. **15**(24-31).

128. Rossmann, M.G., *The locked rotation function.* J Mol Biol, 1972. **64**(1): p. 246-9.

129. Hendrickson, W., *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation.* Science, 1991. **254**: p. 51-8.

130. Perutz, M.F. and I. Ivanyi, *somorphous replacement and phase detrmination in noncentrosymmetric space groups.* Acta Crystallogr A, 1956(9): p. 867-873.

131. Thygesen J, et al., *The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies.* Structure, 1996. **May 15**(4(5)): p. 513-8.

132. S. Banumathi, M.D.Z.D., *Phasing at high resolution using Ta6Br12 cluster.* Acta Cryst., 2003(D59): p. 492-498.

133. Gomis-Ruth, F.X. and M. Coll, *Solving a 300 kDa multimeric protein by low-resolution MAD phasing and averaging/phase extension.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 6): p. 800-5.

134. Lowe, J., et al., *Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution.* Science, 1995. **268**(5210): p. 533-9.

135. Lindqvist, Y., et al., *Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 A resolution.* Embo J, 1992. **11**(7): p. 2373-9.

136. Andersson, I., *Large structures at high resolution: the 1.6 A crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate.* J Mol Biol, 1996. **259**(1): p. 160-74.

137. Andersson, I. and C.I. Branden, *Large single crystals of spinach 1,5-bisphosphate carboxylase/oxygenase suitable for X-ray studies.* J Mol Biol, 1984. **172**(3): p. 363-6.

138. Andersson, I., et al., *Crystallization and preliminary x-ray studies of spinach ribulose 1,5-bisphosphate carboxylase/oxygenase complexed with activator and a transition state analogue.* J Biol Chem, 1983. **258**(23): p. 14088-90.

139. Nar, H., et al., *Atomic structure of GTP cyclohydrolase I.* Structure, 1995. **3**(5): p. 459-66.

140. Schiltz M, et al., *High-pressure krypton gas and statistical heavy-atom refinement: a successful combination of tools for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr., 1997. **53(Pt 1)**: p. 78-92.

141. Soltis MS, S.M., Wiener M, Phillips G, Rees D, J. Appl. Cryst., 1997. **30**: p. 190-194.

142. Naber, N., et al., *A novel adenosine triphosphate analog with a heavy atom to target the nucleotide binding site of proteins.* Protein Sci, 1995. **4**(9): p. 1824-31.

143. Dauter, Z., M. Dauter, and K.R. Rajashankar, *Novel approach to phasing proteins: derivatization by short cryo-soaking with halides.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 2): p. 232-7.

144., *Collaborative Computational Project NThe CCP4 suite: programs for protein crystallography.* ActaCrystallogr, 1994. **D50**: p. 760-3.

145. Rould, M.A., J.J. Perona, and T.A. Steitz, *Improving multiple isomorphous replacement phasing by heavy-atom refinement using solvent-flattened phases.* Acta Crystallogr A, 1992. **48 (Pt 5)**: p. 751-6.

146. RJ, V.F.R., *Noncrystallographic symmetry averaging in phase refinement and extension.* In Methods in Enzymology, 1997. **Vol. CCLXXVII**(Carter CW, Sweet RM (eds)): p. pp 18–53.

147. Zhang, K.Y.J.M., P, *Histogram matching as a new density modification technique for phase refinement and extension of protein molecules.* Acta Cryst., 1990. **A46**(41-46).