APPLYING X-RAY CRYSTALLOGRAPHY FOR ELUCIDATING ENANTIOSELECTIVE ALCOHOL DEHYDROGENASE/CARBONYL REDUCTASE ENZYMES – SUBSTRATE INTERACTIONS

by

TUNG THANH DINH

(Under the Direction of Robert S. Phillips)

ABSTRACT

This dissertation includes four chapters. Chapter 1 includes introduction and literature review. Chapter 2 and chapter 3 were submitted papers. Chapter 4 includes conclusions.

Chapter 2 of this dissertation describes the application of X-Ray crystallography to obtain the tertiary structure of two interesting mutants of secondary alcohol dehydrogenase from *Thermoanaerobacter ethanolicus*, namely C295A and I86A, along with their substrates and cofactor. The focus of the research is to investigate to central role of Zn^{2+} in the catalytic action of the enzyme. This helps to solve the ongoing debate about the existence of pentacoordinated zinc in this thermophilic enzyme and the dynamics of all the residues coordinating to this metal ion. This also sheds a light on the role of Glu60 in the catalysis mechanism of *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase.

Chapter 3 of this dissertation describes the discovery of condition to crystallize the recently found carbonyl reductase from a soil bacterium *Leifsonia xyli*. Applying X-Ray

crystallography results in the first structure of this enzyme at high resolution. A docking program was used to dock BTAP, an important pharmaceutical intermediate, as well as either of the two cofactors NAD(H) and NADP(H). This elucidates the catalytic mechanism of *Leifsonia xyli* carbonyl reductase.

INDEX WORDS: Alcohol dehydrogenase, Biocatalyst, Bioreduction, Carbonyl reductase, Catalytic zinc coordination, Enantioselective reduction, *Leifsonia xyli*, Ligand docking, Protein crystallization, Proton relay system Secondary alcohol dehydrogenase Short-chain alcohol dehydrogenases/reductases (SDRs), *Thermoanaerobacter ethanolicus*, X-Ray crystallography

APPLYING X-RAY CRYSTALLOGRAPHY FOR ELUCIDATING ENANTIOSELECTIVE ALCOHOL DEHYDROGENASE/CARBONYL REDUCTASE ENZYMES – SUBSTRATE INTERACTIONS

by

TUNG THANH DINH

B.S., University of Science Ho Chi Minh city, Vietnam, 2004

M.S., University of Leuven, Belgium, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

APPLYING X-RAY CRYSTALLOGRAPHY FOR ELUCIDATING ENANTIOSELECTIVE ALCOHOL DEHYDROGENASE/CARBONYL REDUCTASE ENZYMES – SUBSTRATE INTERACTIONS

by

TUNG THANH DINH

Major Professor: Robert S. Phillips

Committee: Todd Harrop

Jeffrey Urbauer

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2021

DEDICATION

To my wife and son, my parents and brother for their love, patience, and support over the years

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Prof. Robert S. Phillips, for his guidance throughout my research. He is one of the nicest and most knowledgeable professors I have ever known. He always does his best to help me and treat me with respect. I am fortunate for having him as my advisor. I would also like to thank my advisory committee, Dr. Harrop and Dr. Urbauer, for their guidance and support.

I owe special thanks to my wife for scarifying her career and going to the US with me, my son for being my strength and motivation. I have been forever grateful to my family including my Mom, Dad, and younger brother for their encouragement, patience, and support over the years. I would not be who I am today without them.

I must thank my lab partners: Christopher Nealon and Emma Iradukunda for their suggestions and their help in the lab. I would like to express my thanks Dr. Wood lab members for their training and help for protein crystallization and shipping. I wish to thank the undergraduate students who work with me and contribute greatly to my research. I would like to say thank you to faculty and staff at UGA, especially Ms. Amanda L. Cross and Dr. Richard Hubbard, who have been continuously help students to be as successful as they can be. I have been learning a lot from my professors during the courses I took at UGA. My fellow graduate students in Department of Chemistry also help me tremendously during my time in Athens, GA.

Last but not the least, I would like to show my gratitude to Prof. Li Chun, a visiting scholar from School of Pharmacy, Hangzhou Medical College, Zhejiang, China. He has been cooperating with us in the project of Leifsonia xyli carbonyl reductase and helping tremendously with this expertise and knowledge.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ADHs	Alcohol Dehydrogenases
Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartate
BTAP	[3,5-bis(trifluoromethyl)]acetophenone
BTPE	[3,5- bis(trifluoromethyl)]phenylethanol
C	Celsius
Cys, C	Cystein
E. coli	*Escherichia coli*
ee	Enantiomeric Excess
g	Gram
Gln, Q	Glutamine
Glu, E	Glutamate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His, H	Histidine
HLADH	Horse Liver Alcohol Dehydrogenase
Ile, I	Isoleucine
K	Kelvin
KDa	Kilodalton
L	Liter
L. xyli	*Leifsonia xyli*
Leu, L	Leucine
LXCAR	*Leifsonia xyli* Carbonyl Reductase
Lys, K	Lysine
M	molar
MDRs	Medium-chain Dehydrogenase/Reductases
mg	Milligram
min(s)	Minute(s)
mL	Mililiter
mM	Millimolar
μM	Micromolar
NAD^+	Nicotinamide Adenine Dinucleotide

NADH	Nicotinamide Adenine Dinucleotide, Reduced
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate, Reduced
nm	nanometer
PADH	Primary Alcohol Dehydrogenase
PEGs	Polyethylene glycols
Ph	Phenyl
Phe, F	Phenylalanine
Pro, P	Proline
rpm	round per minute
SADH	Secondary Alcohol Dehydrogenase
SDRs	Short-chain Dehydrogenase/Reductases
SDS-PAGE	Sodium Dodecyl Sulphate–PolyAcrylamide Gel Electrophoresis)
Ser, S	Serine
TbADH	*Thermoanaerobium brockii* Alcohol Dehydrogenase
TeSADH	*Thermoanaerobacter ethanolicus* Secondary Alcohol Dehydrogenase
Thr, T	Threonine
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloride

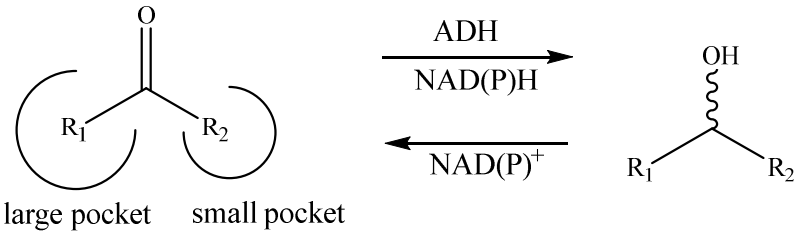
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine
W	Watt
XAS	X-ray Absorption Spectroscopy

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1. Alcohol Dehydrogenase (ADH) family

NAD-dependent [EC 1.1.1.1] and NADP-dependent alcohol dehydrogenases [EC 1.1.1.2], or ADHs, in short, have been proven to be an enzyme class with conserved structures[1]. These enzymes, in the presence of either NADH or NADPH can reduce carbonyls into the corresponding alcohols. In the opposite direction, if the cofactor oxidized forms (NAD^+ or $NADP^+$) are present, alcohols are converted to aldehydes or ketones (Scheme 1.1). The reduction direction brings about either (*R*)- or (*S*)-product, and interestingly, many ADHs are enantioselective, giving rise to mostly either one of these enantiomers. Their enantioselectivity has been explained by a model in which there were two interacting pockets at the active sites, namely the small and large ones. As Prelog's rule suggests[2], the smaller pocket binds the smaller substituent (R_2) from the carbonyl, and vice versa. The relative position of the attacking hydride from the cofactor will decide the product stereochemistry. For instance, if NADPH is on the *re* face of the carbonyl plane, the nucleophilic attack will result in a (*S*)-alcohol (Figure 1.1). One thing should be noted is that here the larger size group (R_1) is assumed to have higher priority than that of R_2 according to the Cahn-Ingold-Prelog rule, which is not always the case. The alcohol absolute configurations must be considered instead of their nomenclature.



Scheme 1.1. General reactions of ADHs

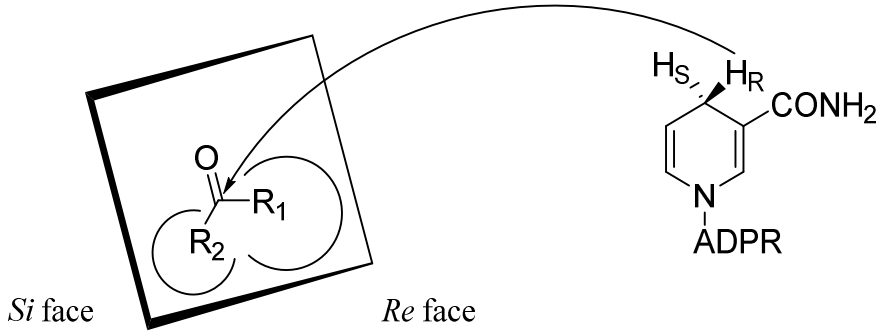


Figure 1.1. Nucleophilic attack of cofactor to carbonyl at ADH active sites [3]

There are several ways to further classify ADHs. For example, some mainly react in the presence of primary alcohols while other mostly reduce secondary alcohol. The former are called primary alcohol dehydrogenase (PADH) and the latter are called secondary alcohol dehydrogenase (SADH), respectively. Another method of categorization is based on the size of the monomer, not the size of the whole enzyme. The first type is short-chain alcohol dehydrogenases/reductases (SDRs), the chain length of which is around 250-300 residues. On the other hand, each monomer of medium-chain alcohol dehydrogenases/reductases (MDRs) is 350-400 amino acid long. Interestingly, many SDRs reduce carbonyls into (*R*)-alcohols while MDRs produce (*S*)-alcohols.

2. *Thermoanaerobacter ethanolicus* Secondary Alcohol Dehydrogenase (TeSADH)

2.1. *Thermoanaerobacter ethanolicus*

A strain (39ET) of thermophilic anaerobacterium was found in thermal waters, muds, and decomposing algal-bacterial mats in Yellowstone National Park. It was isolated and characterized by Lee et al. and named *Clostridium thermohydrosulfuricum* at first. It was afterwards renamed as *Thermoanaerobacter ethanolicus,* strain 39E[4]. Its genome sequence is available under the name *T. ethanolicus* 39E at https://www.ncbi.nlm.nih.gov/nuccore/CP000924.1. While mesophilic yeast species have been utilized in the industry for ethanol product, the processes required separate step of distilling ethanol, as their operating temperatures are not high enough[5]. Thermophilic microorganisms therefore offer a potentially attractive conducting application in which ethanol can be continuously distilled from the cultures at low pressure during fermentation. This would reduce the steps involved and negate the necessity for an ethanol-tolerant organism as the compound evaporated relentlessly from the media. The problem for *T. ethanolicus* is that while being grown in media of 10 g/L glucose or lower, the bacterium essentially produces ethanol, but at higher glucose concentration, acetate and lactase are significant byproducts[6].

2.2. TeSADH

Proteins of *T. ethanolicus* attract some attention due to industrial need of enzymes which can work at high temperature. Bryan[7] and collaborator isolated and characterized two alcohol dehydrogenases from *T. ethanolicus*. The enzyme from JW200 strain prefers primary alcohols[8], while the other is active toward secondary alcohols or ketones in the reverse reaction and, therefore, is referred as *T. ethanolicus* secondary alcohol

dehydrogenase (TeSADH). It was shown to be an NADP dependent Zn^{2+} medium-chain dehydrogenase. The adhB gene which encodes for TeSADH was cloned, sequenced, and expressed[1] in *E.coli*. Amazingly, its amino acid sequence and that of *Thermoanaerobacter brockii* SADH (TbSADH) were different by only 3 residues[9] (Figure 1.2). Our lab resequenced TeSADH and found three errors in the previously published sequence (https://www.uniprot.org/uniprot/P77990), namely Trp-91 → Arg, Pro-313 → Arg, and Gln 325 → Arg. After these three residues are corrected, the TeSADH sequence is 100% identical with that of TbSADH. This proves that the two species are just one but discovered and named independently. Contributing to the confusion is the fact that since 05/31/07, *T. ethanolicus* 39E was renamed as *Thermoanaerobacter pseudoethanolicus* 39E on https://genome.jgi.doe.gov. Because there are three different names for this protein, for the sake of simplicity, only TeSADH is used in this dissertation.

unnamed protein product

Sequence ID: Query_63389 Length: 352 Number of Matches: 1

Range 1: 1 to 352 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
701 bits(1810)	0.0	Compositional matrix adjust.	349/352(99%)	350/352(99%)	0/352(0%)

```
Query  1    MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHE  60
            MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHE
Sbjct  1    MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHE  60

Query  61   AVGEVVEVGSEVKDFKPGDRVVVPAITPDWWTSEVQRGYHQHSGGMLAGWKFSNVKDGVF  120
            AVGEVVEVGSEVKDFKPGDRVVVPAITPDW TSEVQRGYHQHSGGMLAGWKFSNVKDGVF
Sbjct  61   AVGEVVEVGSEVKDFKPGDRVVVPAITPDWRTSEVQRGYHQHSGGMLAGWKFSNVKDGVF  120

Query  121  GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGL  180
            GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGL
Sbjct  121  GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGL  180

Query  181  MAVAGAKLRGAGRIIAVGSRPVCVDAAKYYGATDIVNYKDGPIESQIMNLTEGKGVDAAI  240
            MAVAGAKLRGAGRIIAVGSRPVCVDAAKYYGATDIVNYKDGPIESQIMNLTEGKGVDAAI
Sbjct  181  MAVAGAKLRGAGRIIAVGSRPVCVDAAKYYGATDIVNYKDGPIESQIMNLTEGKGVDAAI  240

Query  241  IAGGNADIMATAVKIVKPGGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRL  300
            IAGGNADIMATAVKIVKPGGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRL
Sbjct  241  IAGGNADIMATAVKIVKPGGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRL  300

Query  301  RMERLIDLVFYKPVDPSKLVTHVFQGFDNIEKAFMLMKDKPKDLIKPVVILA  352
            RMERLIDLVFYK VDPSKLVTHVF+GFDNIEKAFMLMKDKPKDLIKPVVILA
Sbjct  301  RMERLIDLVFYKRVDPSKLVTHVFRGFDNIEKAFMLMKDKPKDLIKPVVILA  352
```

Figure 1.2. Sequence alignment of *T. brockii* SADH and *T. ethanolicus* SADH

TeSADH is a homotetrameric enzyme with molecular weight of 172 kDa and consists of 4 subunits (37.7 kDa), each of which contains one zinc atoms. It is very stable at around 70 °C, with a half-life of two months[1], which is significant improvement from conventional enzyme systems of which stability quickly decreases past 42 °C. The purified protein is even capable of functioning at 90° C or higher with a half-life of about 1.7 hours at 90 °C.[10] More importantly, this protein during preliminary research showed extremely enantio-specificity at high temperature (70 °C)[11].

Chiral alcohols are one of the most important intermediates in pharmaceutical production[12]. Synthesis of only one enantiomer of alcohols has become more and more essential recently[13]. One of the most efficient and popular methods to produce a single enantiospecific alcohol is asymmetric reduction of ketones[14]. Thanks to high stereoselectivity of TeSADH in the reduction direction, which produces only one enantiomeric alcohol from a prochiral ketone, it is a good candidate catalyst for chiral synthesis. The stereospecificity for substrates are decided by structural and chemical properties of the substrate-binding site which contains a large and a small pocket[15]. The current model proposed for the protein stereospecificity is that if the larger substituent of a ketone fits into the large pocket of the active site, the product will be an (*S*)-alcohol assuming that the larger substituent has higher priority than that of the smaller one according to Cahn-Ingold-Prelog rule[16].

The biocatalysts in the industry must possess strong activities and insusceptible to condition changes of the reactions. Temperature and solvent stabilities are essential for a potentially chiral catalyst. Thermozymes, which are enzymes from thermophiles, is therefore more ideal catalysts than mesophilic enzymes[10]. Thermophilic SADHs have been

researched and shown to react with a wide variety of substrates and have high enantioselectivity for prochiral ketone reduction[17]. In fact, there have been several demonstrations of utilizing TeSADH as catalyst in chiral chemistry syntheses[9, 18-20]. However, to popularize the usage of SADHs in general and TeSADH in particular, two major requirements need to be addressed. The first one is the regeneration of the expensive cofactors, especially NADP. In the case of TeSADH, this cofactor can be regenerated by the enzyme itself in the presence of up to 30% (v/v) 2-propanol[21]. Another solution is to mutate the enzyme to change its preference to the less expensive NAD. For instance, Hassler et al. demonstrated that the wild type TeSADH has an apparent K_m (app) value for $NADP^+$ of 0.011 mM, while almost no activity was detected in the presence of NAD or NADH. However, as Tyr218 was changed into phenylalanine by site-directed mutagenesis (Y218F), the mutant showed 5.5-fold higher V_{max} and 2.7-fold lower K_m for NAD^+ than those of the wild type[22]. The second requirement is that enzymes remain stable in the presence of organic solvents and after being immobilized on a solid support. In fact, the industrial application of immobilized yeast and horse liver SADHs (YADH and HLADH) is hindered by their instability at temperature higher than 30 ^oC and susceptibility to organic solvent, which lead to activity loss. On the contrary, these are not issues for TeSADH. Moreover, the immobilized enzyme is extremely stable. After being integrated on Eupergit-C resin, the protein can react with 2-pentanone in 30 days at 37 ^oC with no significant decrease in activity[23].

In addition to the two aforementioned obstacles, the thermophiles grow slower and yield lower cell densities, which necessitate the search for other expression system. Instead of using *T. ethanolicus* directly in a whole-cell bioreduction method, expression in *E. coli*

has been utilized and proved to be efficient, producing good yield of TESADH[1]. This approach is advantageous as it does not require high temperature to grow *T. ethanolicus* anaerobically. Therefore, *E. coli* can be grown in large scale conveniently and produce great amount of recombinant TeSADH. Thanks to its thermal stability, the purification process can be simplified by incubating of the cell-free extract at high temperature. This leads to the precipitation of most denatured *E. coli* proteins.

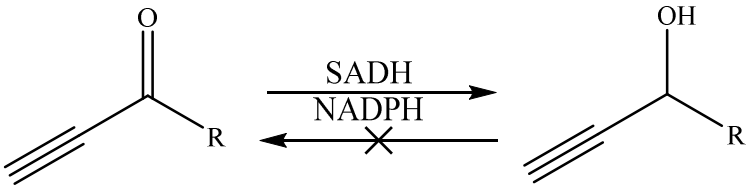
Standard activity and kinetic assays

The standard TeSADH activity is defined as $NADP^+$ reduction in the presence of a substrate at 60 ^oC as described by Burdette et al[24]. Prior to activity being measured, the enzyme solution was incubated at 55 ^oC for 15 minutes. The apparent values for K_m (app) and V_{max} (app) were determined by varying substrate concentration from 0.2 K_m (app) to 20 K_m (app). Those values for $NADP^+$ was obtained by experiment with propanol-2 concentration of 20 K_m (app) at 60 ^oC[1]. Kinetic assay of various substrates showed that TeSADH has very weak activity toward primary alcohols. More specifically, the k_{cat}/K_m as well as K_m values for primary alcohols were about 50-fold higher and 170-fold lower those for secondary counterpart, respectively[1].

2.3. TeSADH C295A mutant

In this project, we have been working on two TeSADH mutants: C295A and I86A. As the wild type enzyme reacts with alkyl ethynyl ketones, it is inactivated (Scheme 1.2). It was first hypothesized that the ethynyl group was attacked by Cys295 in a Michael addition fashion, leading to inactivation (Figure 1.3). While the C295A mutant is still inactivated, which invalidates the hypothesis, it showed an interesting enantioselectivity compared to the wild type enyzme. The wild type, when reacting with various alkyl ethynyl

ketones, reverses its enantioselectivity if the alkyl group is isobutyl or larger, producing (*R*) instead of (*S*) alcohol. In contrast, the C295A mutant retains its enantioselectivity unless a sterically massive alkyl group was present[15]. This was explained by the fact that the alkyl group, counterintuitively, preferentially bound to the small pocket because this pocket is more hydrophobic that the big one. If alkyl group gets too big, it can only be accommodated by the bigger pocket in the wild type, resulting in the enantiospecific reversal. Thus, by replacing the Cys295 with alanine, the smaller pocket is enlarged and can contain a larger alkyl group, resulting in the enantioselectivity retention of the enzyme.



Scheme 1.2. Conversion of Alkyl ethynyl ketone by TeSADH

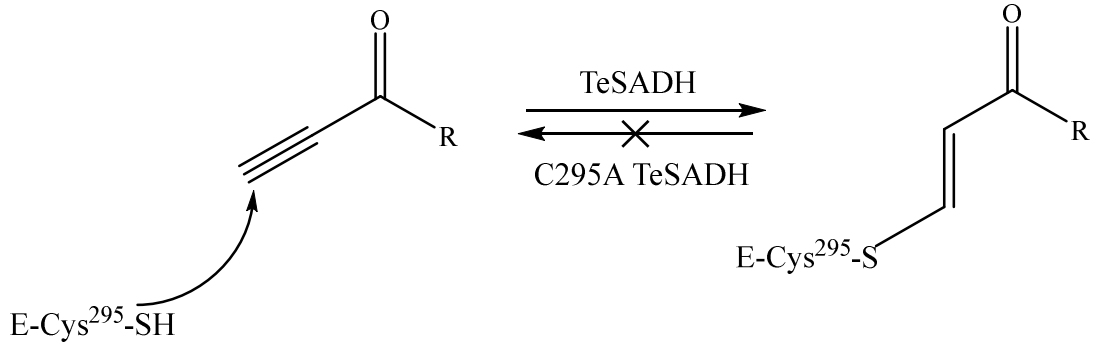
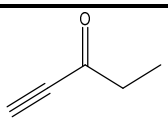
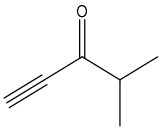
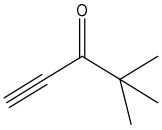
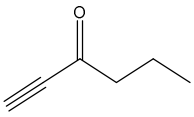
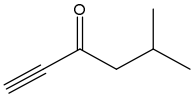


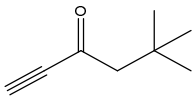
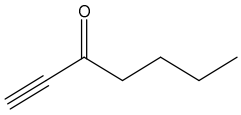
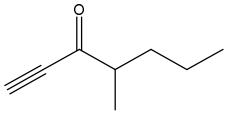
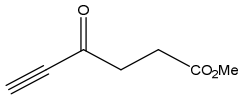
Figure 1.3. Hypothetical Michael addition between Cys-295 and alkyl ethynyl ketone

Yields, absolute configurations, and enantiomeric excess (ee) from the reduction of alkyl ethynyl ketones are summarized in Table 1.1. When the alkyl size is increased from ethyl to isopropyl (entry 1 and 2), the reaction yield and ee value for C295A went up significantly compared to those of the wild-type. However, if the alkyl group is *tert*-butyl,

while C295A conversion yield drops, the wild-type has no activity toward the substrate. These implies that the orientation of the carbon α determines the hydrophobic interaction strength. While the α-branching may bring about better substrate interaction involving in hydrophobic residues at the active site, too much of it results in steric tension. As the size of the alkyl group increases (from entry 4-6), the reaction yields and ee gradually decrease for both the wild type and C295A. The β branching was also demonstrated to give a negative impact on substrate binding. Most importantly, the wildtype stereospecificity is reversed when alkyl group is isobutyl and onward, whereas that of C295A is retained (entry 5-8) until the alkyl was replaced by an extremely large group like methyl propionate (entry 9). Another important observation is that the ethynyl *sec*-pentyl ketone led to very high ee though low reaction yield, which further proved the significance of α-branching.

Table 1.1. Reduction of alkyl ethynyl ketones by TeSADH- wildtype and C295A[15]

Entry	Substrate	Yield [a,b] (%)	Abs. conf [a]	ee [a] (%)
1		39 (32)	S (S)	76 (80)
2		88 (50)	S (S)	>98 (>98)
3		39 (0)	S (S)	85 (85)
4		51 (28)	S (S)	76 (51)
5		42 (20)	S (R)	56 (50)

6		0 (0)	—(R)	—(66)
7		60 (32)	S (R)	67 (42)
8		43 (0)	S (–)	>98 (–)
9		23 (35)	R (R)	60 (82)

[a] Absolute configuration. Results of wild-type are shown in parentheses.

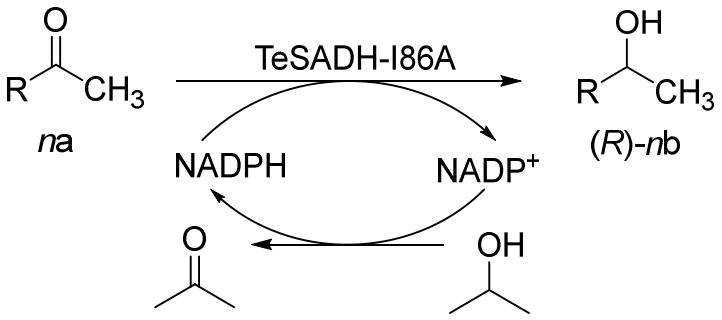
[b] Yields shown are isolated yields.

2.4. TeSADH I86A mutant

The other mutant, TeSADH-I86A, can catalyze the reaction of aromatic ketones, a characteristic the wild type doesn't have[3]. Similar to C295A mutant, the smaller pocket is extended such that it can contain the phenyl group from aromatic ketones. As a result, the reduction of alkyl phenyl ketone will produce (*R*)-alcohols. Furthermore, this specificity is expanded to benzylic and heteroaryl ketones which are not substrate for the wild type enzyme. The optimal pH for TeSADH-I86A reduction and oxidation are 6.5 and 8, respectively. When reacting with acetophenone and (*R*)-1-phenylethanol, I86A gave the K_m (app) of 0.61 and 0.74 mM, and the corresponding V_{max} values are 18.0 and 32.2 μM/min/mg, respectively. On the contrary, the wild type protein showed no detectable activity toward these two substrates. However, the first downside of this mutant is that it showed weaker affinity and lower activity toward aromatic ketones compared to their aliphatic counterpart. Moreover, since acetophenone is poorly soluble in aqueous solution, this necessitates the usage of mediating solvent or a solvent system like acetonitrile and water mixture (Table 1.2).

Table 1.2. Kinetic parameters of TeSADH-I86A in reaction with acetophenone, (*R*)-1-phenylethanol, and (*S*)-1-phenylethanol[3]

Substrate	K_m (app) (mM)	V_{max} (μmol/min/mg)	V_{max}/K_m (app) (mL/min/mg)
Acetophenone	0.61 ± 0.134	18 ± 1.36	0.031
(*R*)-1-phenylethanol	0.74 ± 0.045	32.2 ± 0.64	0.61 ± 0.134
(*S*)-1-phenylethanol	-	-	-



Scheme 1.3. Reduction of aryl methyl ketones

Kinetic assays for methyl ketones of several phenyl and heteroaryl substituents (Scheme 1.3) were conducted at 50 ºC in 1 mL 50 mM Tris-HCl buffer (pH 8.0 adjusted at 25 ºC) with TeSADH-I86A (0.53 mg), $NADP^+$ (2.0 mg), substrates (0.42 mmol), and 2-propanol (400 μL) which is utilized as co-substrate to regenerate NADPH. Except for substrate 1a, buffers were supplemented with $ZnCl_2$ (10 μM), DTT (2.0 mg) since this addition was proven to improve reaction yield from 47% to 79% in case of acetophenone (1a and 2a). In contrast to the activity assay, biphasic media of hexane:water was used instead of mediating solvent. From Table 1.3, it can be clearly seen that all reactions resulted in high enantioselectivity (98-99% ee) but with varying conversion rates. For instance, conversion for acetophenone (2a) is 79% but drops to 60% for propiophenone

(not listed in Table 1.3). The mutant however showed no detectable activity toward butyrophenone due to its poor solubility in aqueous solution. This also happened to substituted phenyl substrates, including *2*-methyl-, 4-methyl-, 4-hydroxy, 4-chloro-, or 3-methoxy acetophenone, which implies that the extended large pocket is still not big enough to accommodate a more sterically demanding substituted phenyl ring. The only exception is 2,4-difluoroactephone which is converted at 60%, likely due to the small size of the two fluorine groups. This issue might be circumvented in the future by mutating more residues in the small pocket to make it even bigger. For heteroaryl ketones, 3-substituted rings gave rise to average to low yield (5a and 7a) while 2- and 4- substituted rings led to almost complete conversion, which can be explained by the inductive effect. These results offer very promising pharmaceutical applications in the future, for instance, product 4b is used in the synthesis of PNU-142721, an AIDS drug candidate[25]. Compared to the wild type enzyme which produces (*S*)-alcohols during reduction, I86A TeSADH generates the (*R*)-alcohols, as the phenyl groups, which generally have higher priority according to Cahn-Ingold-Prelog rule, fit into the small pocket. As a result, the hydride from the cofactor will be on the *si* face of the ketones instead of *re* face as in the wildtype.

Table 1.3. Reduction of aryl methyl ketones by TeSADH-I86A

Substrate	R	Product [a]	Conversions [%] [b]	ee [%] [c]
1a	Ph	(*R*)-1b	47 [d]	98
2a	Ph	(*R*)-2b	79	98
3a	2,4-Difluorophenyl	(*R*)-3b	33	> 99
4a	2-Pyridyl	(*R*)-4b	> 99	> 99
5a	3-Pyridyl	(*R*)-5b	46	> 99

6a	4-Pyridyl	(*R*)-6b	> 99	> 99
7a	3-Thienyl	(*R*)-7b	76	> 99

[a] The absolute configurations were determined by comparing optical rotation values with references, [b] % Conversions were calculated by GC data, [c] ee values were computed by chiral stationary phase GC for the corresponding acetate derivatives, [d] assay was performed in the absence of DTT and $ZnCl_2$.

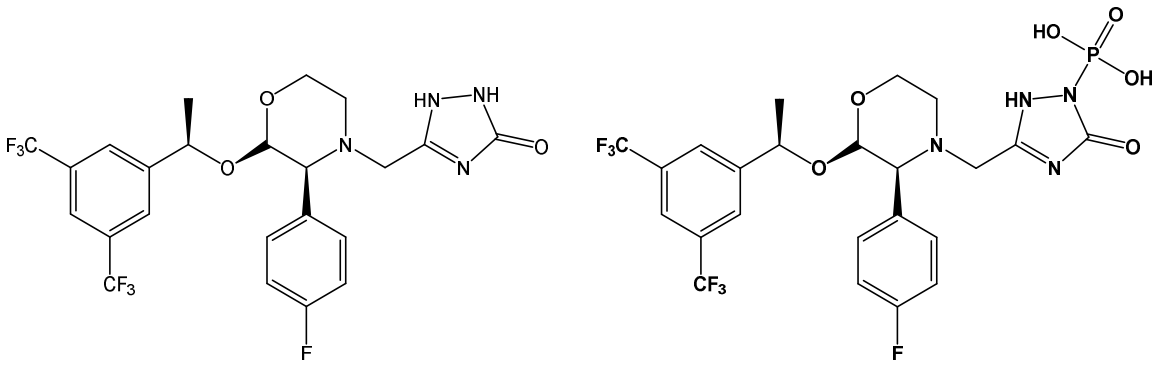
In this research project, we plan to apply X-ray crystallography to obtain the structure of the TeSADH mutants, including C295A and I86A, along with their substrates and cofactor for addressing several scientific questions: 1) What is the role of Zn^{2+} in the catalytic activity of the enzyme; 2) What is the dynamics of residues coordinating to Zn^{2+} and their effects on TeSADH catalysis; and 3) What are the dynamics of amino acid interaction with the cofactor.

3. Leifsonia xyli carbonyl reductase (LXCAR)

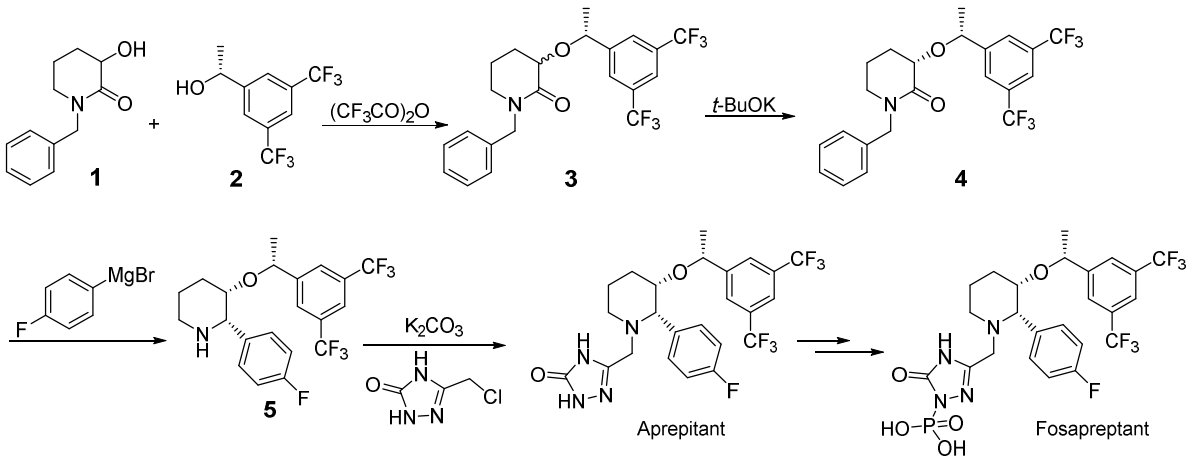
3.1. Significance and syntheses of Aprepitant and Fosaprepitant

Aprepitant (commercial name: Emend) and Fosaprepitant (commercial name: Ivemend) are popular drugs (Scheme 1.4) which have been described for treating side effect of chemotherapy during cancer treatment, including nausea and vomiting[26-28]. Fosaprepitant is merely a pro-drug of Aprepitant in which a phosphate group is added to the latter‘s triazole ring; therefore their syntheses are almost the same (Scheme 1.5)[29]. The synthesis pathway requires the key chiral compound (1*R*)-[3,5-bis(trifluoromethyl)]phenylethanol or (*R*)- BTPE (intermediate 2). While BTPE can be produced by reducing [3,5-bis(trifluoromethyl)]acetophenone or BTAP (Scheme 1.6) with reducing reagent like lithium aluminium hydride ($LiAlH_4$) or sodium borohydride

($NaBH_4$), the issue is that these reagents will lead to a racemic mixture. The following enantiomer separation is painstaking and expensive; hence, it can be said that (*R*)- BTPE is the bottleneck of Aprepitant and Fosaprepitant syntheses. Since the demand for the two drugs has been increasing, the need of efficient and cost-effective production of (*R*)- BTPE has become more and more significant and has been addressed[30-31].

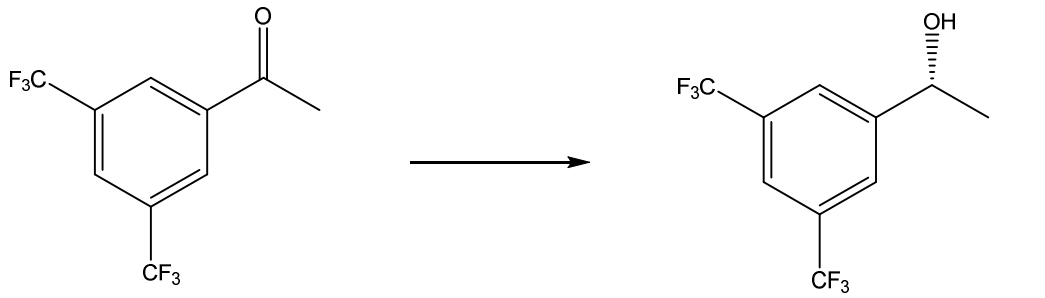


Scheme 1.4. Structures of Aprepitant and Fosaprepitant



Scheme 1.5. The synthesis pathway of Aprepitant and Fosaprepitant [29]

3.2. Enantioselective bioreduction of BTAP



3,5-bis(trifluoromethyl)acetophenone (BTAP) (*R*)-[3,5-bis(trifluoromethyl)phenyl] ethanol (BTPE)

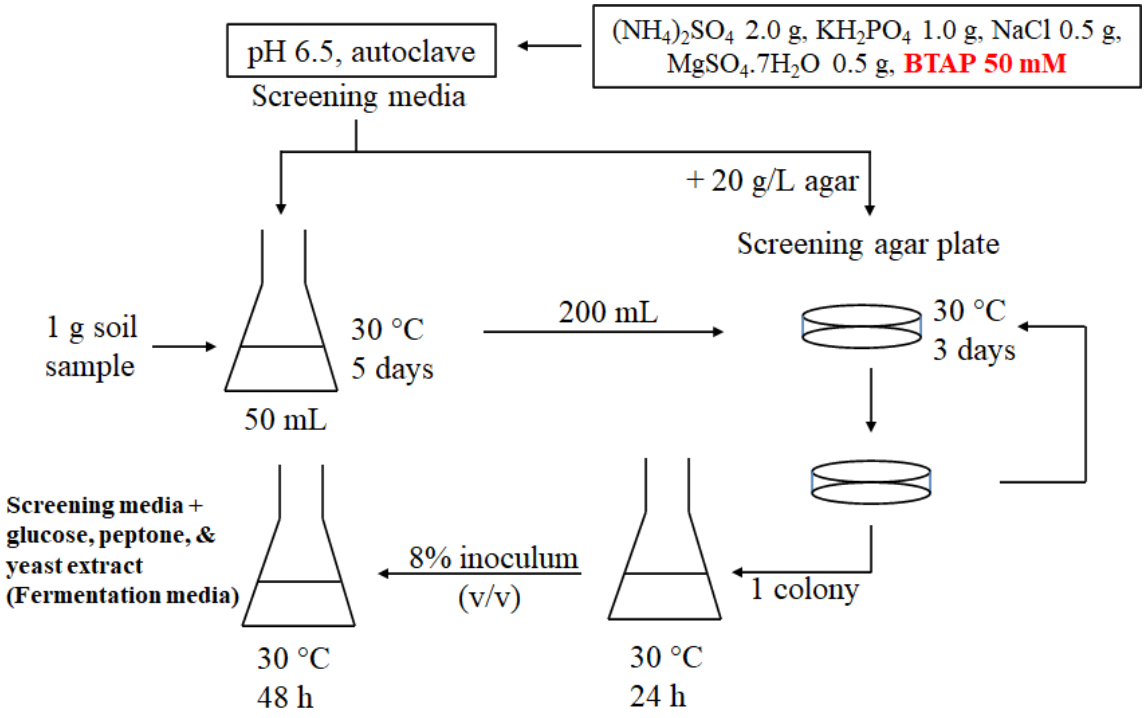
Scheme 1.6. Reduction of BTAP into (*R*)-BTPE

Scientists discovered several chemical methods for the synthesis of (*R*)- BTPE by BTAP reduction[32-34]. Still, other approaches of bioreduction has recently been reported by numerous groups in which various microbial species were found to be capable of reducing BTAP to (*R*)-BTPE. For example, Gelo-Pujic et al.[35] reduced BTAP with *Lactobacillus kefir*, achieving over 99% ee, though only at very low substrate concentration (no more than 5 mM). Moreover, a strain of *Penicillium expansum* isolated from soil sample in Erzurum, Turkey led to 99% ee at moderate yield (76%) but requiring long reaction time (56 h)[36].

3.3. Discovery and bioreduction of LXCAR

In 2011, in an attempt to find microorganisms able to selectively reduce BTAP to (*R*)-BTPE, a project was subsidized to screen soil microbes in China for this reduction. The screening was detailed by Wang et al.[37] but in short the sample, including bacteria and fungi, was cultured in a BTAP-containing screening media consisting of several inorganic salts (Scheme 1.7). The media were prepared by adjusting to pH 6.5 and autoclaved before the soil sample was added and shaken for 5 days. Subsequently, 0.2 mL samples were

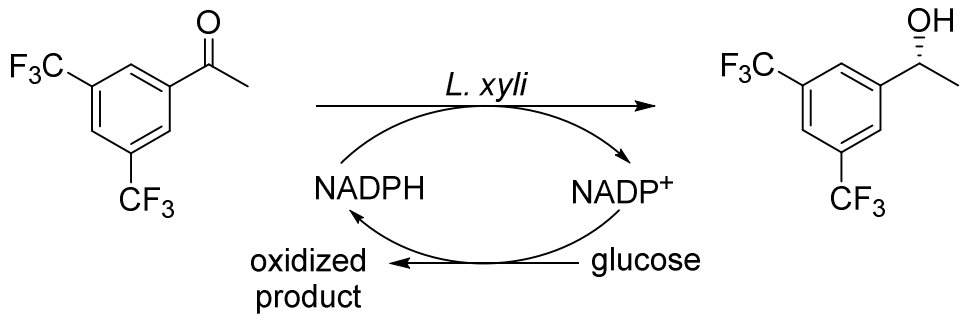
plated on agar media which is simply the liquid screening media supplemented by agar (20 g/L). The microorganism strains that could survive with BTAP being the only carbon-source resulted in lawn of colonies. They were separated by continuous streaking on fresh agar plate until single colonies could be isolated. Each colony subsequently was incubated again in screening media before being amplified by fermentation, which was the same as the former, but supplemented with glucose, peptone, and yeast extract. Finally, dry cells of a strain were obtained by centrifugation and washing twice with NaCl aqueous solution for further investigation and storage. All incubation were done at 30 °C with autoclaved media.



Scheme 1.7. Screening of soil sample leading to the discovery of *L. xyli*

The screening resulted in hundreds of strains which were then tested for their ability to reduce BTAP into (*R*)-BTPE. Among these strains, HS0904 worked the best on BTAP[37]. This strain was later identified as *Leifsonia xyli* (or *L. xyli*). The preliminary whole-cell

bioreduction of it was conducted with 2 g wet cells in a 10 mL reaction mixture of 200 mM phosphate buffer (pH 8.0), 0.5 mM BTAP, and 1 g of glucose as co-substrate for NADPH regeneration (Scheme 1.8). The reaction flask was shaken at 200 rpm in 24 hours at 30 ^{o}C, and the reaction was stopped by removing the biomass with centrifugation. The supernatant was then extracted with ethyl acetate and the resulting extract was finally analyzed by gas chromatography (GC). Later, Wang et al. tried to investigate the best condition for this bioreduction by performing multiple reactions with varying factors such as buffer, pH, temperature, shaking speed and BTAP concentration. The best condition gave good yields and enantiomeric excess values[38] after 30 hours at 62% and 99.4%, respectively starting with relatively high concentration of substrate (70 mM). Per dry cell weight, the carbonyl reductase activity was at 1120 unit/g. An improvement of the whole-cell approach was discovered by Ouyang et al.[39] The group performed BTAP bioreduction in the presence of different carbohydrates and alcohols as cosubtrates. Among them, isopropanol gave best result of 91.8 % yield and 99.9% ee with higher BTAP concentration 200 mM. However, this substrate concentration is still not high enough for industrial scale. Thus, development of more robust mutants was required to enhance the catalytic efficiency.



Scheme 1.8. BTAP bioreduction by *L. xyli*

3.4. LXCAR enzyme

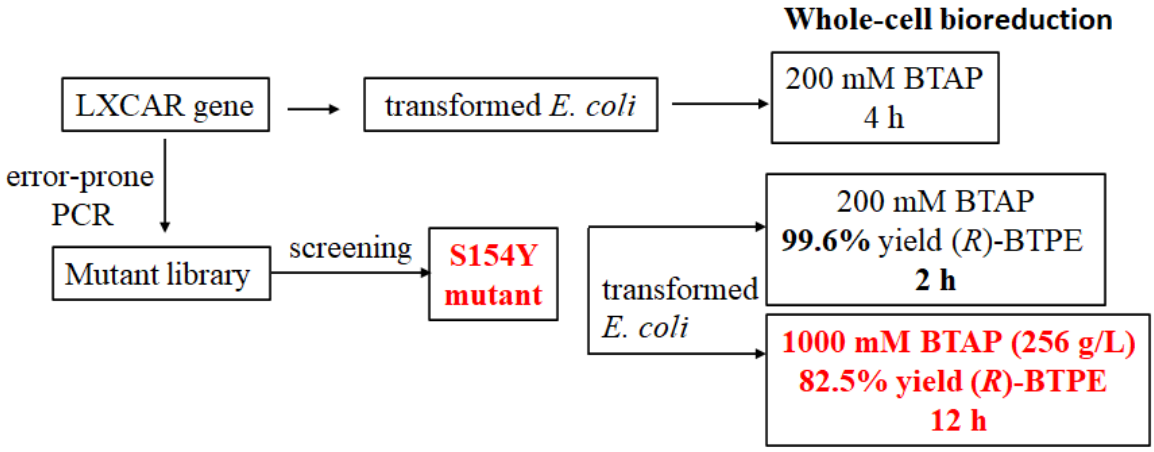
Bioinformatic analysis of the *Leifsonia xyli* genome revealed that a carbonyl reductase (CAR) gene was responsible for its ability to convert BTAP into (*R*)-BTPE. Analyzing the N-terminal amino acid sequence led the conclusion that the LXCAR protein from *L. xyli* is part of the SDRs family. Moreover, it was demonstrated by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) to be a homodimer consisting of two identical chains of 24 kDa[40]. The process of cloning, expression, and purification of LXCAR were detailed by Wang et al.[40] This group also measured the protein activities and stabilities, varying different elements like buffer type, pH, temperature, presences of various metal ions, and most importantly, substrates[40]. LXCAR was shown to be the most stable and active at pH 7-7.2 and 34 ^{o}C. Interestingly, several ions like Li^{+}, Ca^{2+}, Mg^{2+}, Mn^{2+} and Co^{2+} were demonstrated to help to increase the enzyme activity, especially Mg^{2+} and Mn^{2+}. Other metal ions had almost no effect on LXCAR activity except for Cu^{2+}, Sn^{2+}, Fe^{2+}, Pb^{2+} and Fe^{3+} which significantly decrease the ezyme activity. In term of substrate specificity, enzymatic activities were measured in the presence of numerous substrates including ketones, keto esters, and aldehydes and then comparing with that of BTAP which was considered as standard (100%). The results demonstrated that while possessing a relatively broad specificity, the LXCAR shows very weak activity toward aliphatic aldehyde and ketones such as acetone, 2-butanone, and 4-methyl-2-pentanone. But the most striking result was that LXCAR was very active toward monosubstituted trifluoromethylacetophenone (131%) but it has little to no activity toward acetophenone and a number of its mono- or disubstituted halogen derivatives including 2'-chloroacetophenone, 4'-fluoroacetophenone, and 2',6'-dichloro-3'-fluoroacetophenone.

The only groups that led to significant activity were β-hydroxyketone and β-diketones. For example, 4-hydroxy-2-butanone and 2,4-pentanedione gave rise to approximately the same and half as much activity as that of BTAP, respectively. β-Ketoesters in some cases brought about even higher activity, for instance, achieving 143% and 240% relative activity for *tert*-butyl (*R*)-6-cyano-5-hydroxyl-3-oxohexanoate and ethyl acetoacetate, respectively. A very attracting characteristic of this protein is that it can work with either NAD(H) or NADP(H) as a cofactor with similar K_m and k_{cat}, while many alcohol dehydrogenases (ADHs) are dependent on more expensive NADP(H) as cofactor. The fact that LXCAR can react with less expensive NAD(H) makes it even more promising to be utilized as biocatalyst in cost-effective manufacturing.

3.5. LXCAR-S154Y mutant

The industrial-scale application of biocatalysts requires high substrate concentration of at least 100 g/L and conversion rate above 95% [41-42]. Wang et al.[38] expressed LXCAR gene in *Escherichia coli* (*E. coli*) and applied cells to BTAP bioreduction. The reaction was improved markedly, the time of which decreased from 30 hours of the standard process (mentioned above in section 3.3) to 4 hours. Furthermore, the crude extract of the expressed cells exhibited 62-fold as much activity as that of *L. xyli*, reaching 1.54 U/mg. One unit (U) was defined as the amount of LXCAR that produces 1 μg of BTPE from BTAP per minute. Subsequently, Wang et al.[38] established a mutant library by random mutation with error-prone PCR (epPCR). About 500 transformants were screened for BTAP reduction activities by measuring absorbance decrease of NADH at 340 nm in the presence of this substrate. Five of them which exhibited higher activities higher than that of the wild type enzyme were then singled out by agar plate, amplified in

cultures, lysed, and centrifuged to obtain crude extracts. After that, the activities of these extracts were determined again by NADH absorbance while the conversion rates and ee values were analyzed by GC as mentioned above in section 3.3. By comparing these results with that of the wild type crude extract, the best transformant with satisfactory reaction yield was chosen and DNA-sequenced. Aligning DNA sequences revealed the four nucleotide bases that were changed were C461A, G462T, C580A, and C682T, corresponding to two mutation sites: serine to tyrosine at position 154 and leucine to isoleucine at position 194. The former position is part of the $SX_{12}YX_3K$ motif that is demonstrated to be essential for the catalytic activity of SDRs family.



Scheme 1.9. Bioreduction of *E. coli* cells transformed with wild type and S154Y mutant LXCAR genes.

To assess the effect of each residue substitution, Wang et al. generated two single mutants (S154Y and L94I) and one double mutant. Their crude extracts along with that of the double mutants were then subjected to asymmetric reduction of BTAP. Interestingly, the S154Y had better activity than the double mutant (S154Y/L194I) and the other single mutant L194I. The specific activities of purified LXCAR-S154Y/L194I and LXCAR-

S154Y toward BTAP were 23.4 and 18.6 U/mg, respectively. When the transformed *E. coli* cells of LXCAR-S154Y was subjected to whole-cell bioreduction technology (Scheme 1.9) in the presence of 200 mM BTAP, this gave rise to 99.6% conversion rate and 99% ee after only 2 hours, compared to 4 hours of the wild type. More impressively, as BTAP concentration was increased to 1 M, S154Y *E. coli* cells converted 82.5% of this amount after 12 hours. This is so far the highest substrate concentration and conversion rate reported for the enantioselective reduction of BTAP to (*R*)-BTPE. Application wise, compared to the chemistry approach, biocatalysis offers some advantages including milder reaction conditions, less byproduct, high enantioselectivity, no residual metals in the product, and negating the need of protection and deprotection during reactions. As a result, the biocatalytic method is generally not only more efficient but also more environmentally friendly[43]. In this research project, we plan to employ X-ray crystallography to obtain the structure of the LXCAR-S154Y mutant to address several scientific questions: 1) What are the interactions of the enzyme and BTAP at the active site; 2) Why this mutant reacts better with BTAP; 3) Which interactions are there between the enzyme and cofactor, either NAD(H) or NADP(H); and 4) Why this enzyme can use either cofactor?

References

1. Burdette, D. S.; Vieille, C.; Zeikus, J. G., Cloning and expression of the gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. *Biochemical Journal* **1996,** *316* (1), 115-122.

2. Prelog, V., Specification of the stereospecificity of some oxido-reductases by diamond lattice sections. *Pure Appl. Chem* **1964,** *9* (1), 119-130.

3. Musa, M. M.; Lott, N.; Laivenieks, M.; Watanabe, L.; Vieille, C.; Phillips, R. S., A single point mutation reverses the enantiopreference of *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase. *ChemCatChem* **2009,** *1* (1), 89-93.

4. Lee, Y.-E.; Jain, M. K.; Lee, C.; Zeikus, J. G., Taxonomic distinction of saccharolytic thermophilic anaerobes: description of *Thermoanaerobacterium xylanolyticum* gen. nov., sp. nov., and *Thermoanaerobacterium saccharolyticum* gen. nov., sp. nov.; reclassification of *Thermoanaerobium brockii*, *Clostridium thermosulfurogenes*, and *Clostridium thermohydrosulfuricum* E100-69 as *Thermoanaerobacter brockii* comb. nov., *Thermoanaerobacterium thermosulfurigenes* comb. nov., and *Thermoanaerobacter thermohydrosulfuricus* comb. nov., respectively; and transfer of *Clostridium thermohydrosulfuricum* 39E to Thermoanaerobacter ethanolicus. *International Journal of Systematic and Evolutionary Microbiology* **1993,** *43* (1), 41-51.

5. Benda, I.; Reed, G., Prescott and Dunn's Industrial Microbiology. **1982**.

6. Bryant, F.; Ljungdahl, L. G., Characterization of an alcohol dehydrogenase from *Thermoanaerobacter ethanolicus* active with ethanol and secondary alcohols. *Biochemical and biophysical research communications* **1981,** *100* (2), 793-799.

7. Bryant, F. O.; Wiegel, J.; Ljungdahl, L. G., Purification and properties of primary and secondary alcohol dehydrogenases from *Thermoanaerobacter ethanolicus*. *Applied and environmental microbiology* **1988,** *54* (2), 460-465.

8. Wagner, I. D.; Wiegel, J., Diversity of thermophilic anaerobes. *Annals of the New York Academy of Sciences* **2008,** *1125* (1), 1-43.

9. Peretz, M.; Burstein, Y., Amino acid sequence of alcohol dehydrogenase from the thermophilic bacterium *Thermoanaerobium brockii*. *Biochemistry* **1989,** *28* (16), 6549-6555.

10. Burdette, D. S.; Tchernajencko, V.; Zeikus, J. G., Effect of thermal and chemical denaturants on *Thermoanaerobacter ethanolicus* secondary-alcohol dehydrogenase stability and activity. *Enzyme and microbial technology* **2000,** *27* (1-2), 11-18.

11. Pham, V. T.; Phillips, R. S., Effects of substrate structure and temperature on the stereospecificity of secondary alcohol dehydrogenase from *Thermoanaerobacter ethanolicus*. *Journal of the American Chemical Society* **1990,** *112* (9), 3629-3632.

12. Genov, D. G.; Ager, D. J., Asymmetric hydrogenation of ketones catalyzed by Ru^{II}–bicp complexes. *Angewandte Chemie* **2004,** *116* (21), 2876-2879.

13. Kurbanoglu, E. B.; Zilbeyaz, K.; Ozdal, M.; Taskin, M.; Kurbanoglu, N. I., Asymmetric reduction of substituted acetophenones using once immobilized *Rhodotorula glutinis* cells. *Bioresource technology* **2010,** *101* (11), 3825-3829.

14. Matsuda, T.; Yamanaka, R.; Nakamura, K., Recent progress in biocatalysis for asymmetric oxidation and reduction. *Tetrahedron: Asymmetry* **2009,** *20* (5), 513-557.

15. Heiss, C.; Laivenieks, M.; Zeikus, J. G.; Phillips, R. S., Mutation of cysteine-295 to alanine in secondary alcohol dehydrogenase from *Thermoanaerobacter ethanolicus* affects

the enantioselectivity and substrate specificity of ketone reductions. *Bioorganic & medicinal chemistry* **2001,** *9* (7), 1659-1666.

16. Prelog, V., Specification of the stereospecificity of some oxido-reductases by diamond lattice sections. *Pure and Applied Chemistry* **1964,** *9* (1), 119-130.

17. Keinan, E.; Hafeli, E. K.; Seth, K. K.; Lamed, R., Thermostable enzymes in organic synthesis. 2. Asymmetric reduction of ketones with alcohol dehydrogenase from *Thermoanaerobium brockii*. *Journal of the American Chemical Society* **1986,** *108* (1), 162-169.

18. Keinan, E.; Seth, K. K.; Lamed, R., Organic synthesis with enzymes. 3. TBADH-catalyzed reduction of chloro ketones. Total synthesis of (+)-(*S*,*S*)-(*cis*-6-methyltetrahydropyran-2-yl)acetic acid: a civet constituent. *Journal of the American Chemical Society* **1986,** *108* (12), 3474-3480.

19. Keinan, E.; Seth, K. K.; Lamed, R.; Ghirlando, R.; Singh, S. P., Thermostable Enzymes in Organic Synthesis, 4. Tbadh-Catalyzed Preparation of Bifunctional Chirons. Total Synthesis of *S*-(+)-*Z*-Tetradec-5-En-13-Olide. *Biocatalysis* **1990,** *3* (1-2), 57-71.

20. Lamed, R.; Keinan, E.; Zeikus, J., Potential applications of an alcohol-aldehyde/ketone oxidoreductase from thermophilic bacteria. *Enzyme and Microbial Technology* **1981,** *3* (2), 144-148.

21. Hummel, W.; KULA, M. R., Dehydrogenases for the synthesis of chiral compounds. *European Journal of Biochemistry* **1989,** *184* (1), 1-13.

22. Hassler, B. L.; Dennis, M.; Laivenieks, M.; Zeikus, J. G.; Worden, R. M., Mutation of Tyr-218 to Phe in *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase:

Effects on bioelectronic interface performance. *Applied biochemistry and biotechnology* **2007,** *143* (1), 1-15.

23. Keinan, E.; Seth, K. K.; Lamed, R., Synthetic applications of alcohol-dehydrogenase from *Thermoanaerobium brockii*. *Annals of the New York Academy of Sciences* **1987,** *501*, 130-149.

24. Burdette, D.; Zeikus, J., Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2 Adh) as a bifunctional alcohol dehydrogenase-acetyl-CoA reductive thioesterase. *Biochemical Journal* **1994,** *302* (1), 163-170.

25. Stampfer, W.; Edegger, K.; Kosjek, B.; Faber, K.; Kroutil, W., Simple biocatalytic access to enantiopure (*S*)-1-heteroarylethanols employing a microbial hydrogen transfer reaction. *Advanced Synthesis & Catalysis* **2004,** *346* (1), 57-62.

26. Jin, Y.; Wu, X.; Guan, Y.; Gu, D.; Shen, Y.; Xu, Z.; Wei, X.; Chen, J., Efficacy and safety of aprepitant in the prevention of chemotherapy-induced nausea and vomiting: a pooled analysis. *Supportive Care in Cancer* **2012,** *20* (8), 1815-1822.

27. Wu, C.-E.; Liaw, C.-C., Using aprepitant as secondary antiemetic prophylaxis for cancer patients with cisplatin-induced emesis. *Supportive Care in Cancer* **2012,** *20* (10), 2357-2361.

28. Nakade, S.; Ohno, T.; Kitagawa, J.; Hashimoto, Y.; Katayama, M.; Awata, H.; Kodama, Y.; Miyata, Y., Population pharmacokinetics of aprepitant and dexamethasone in the prevention of chemotherapy-induced nausea and vomiting. *Cancer chemotherapy and pharmacology* **2008,** *63* (1), 75-83.

29. Wan, W.-L.; He, Y.; Guan, M.; Li, X.-L.; Cheng, X.; Wu, Y., Synthesis of the major isomers of Aprepitant and Fosaprepitant. *Chinese Chemical Letters* **2013,** *24* (12), 1118-1122.

30. Zhao, M. M.; McNamara, J. M.; Ho, G.-J.; Emerson, K. M.; Song, Z. J.; Tschaen, D. M.; Brands, K. M.; Dolling, U.-H.; Grabowski, E. J.; Reider, P. J., Practical asymmetric synthesis of aprepitant, a potent human NK-1 receptor antagonist, via a stereoselective Lewis acid-catalyzed trans acetalization reaction. *The Journal of organic chemistry* **2002,** *67* (19), 6743-6747.

31. Brands, K. M.; Payack, J. F.; Rosen, J. D.; Nelson, T. D.; Candelario, A.; Huffman, M. A.; Zhao, M. M.; Li, J.; Craig, B.; Song, Z. J., Efficient synthesis of NK1 receptor antagonist aprepitant using a crystallization-induced diastereoselective transformation. *Journal of the American Chemical Society* **2003,** *125* (8), 2129-2135.

32. Hansen, K. B.; Chilenski, J. R.; Desmond, R.; Devine, P. N.; Grabowski, E. J.; Heid, R.; Kubryk, M.; Mathre, D. J.; Varsolona, R., Scalable, efficient process for the synthesis of (*R*)-3,5-bistrifluoromethylphenylethanol via catalytic asymmetric transfer hydrogenation and isolation as a DABCO inclusion complex. *Tetrahedron: Asymmetry* **2003,** *14* (22), 3581-3587.

33. Pollard, D.; Truppo, M.; Pollard, J.; Chen, C.-y.; Moore, J., Effective synthesis of (*S*)-3,5-bistrifluoromethylphenylethanol by asymmetric enzymatic reduction. *Tetrahedron: Asymmetry* **2006,** *17* (4), 554-559.

34. Li, W.; Sun, X.; Zhou, L.; Hou, G.; Yu, S.; Zhang, X., Highly Efficient and Highly Enantioselective Asymmetric Hydrogenation of Ketones with TunesPhos/1,2-Diamine−Ruthenium (II) Complexes. *The Journal of organic chemistry* **2009,** *74* (3), 1397-1399.

35. Gelo-Pujic, M.; Le Guyader, F.; Schlama, T., Microbial and homogenous asymmetric catalysis in the reduction of 1-[3,5-bis(trifluoromethyl)phenyl]ethanone. *Tetrahedron: Asymmetry* **2006,** *17* (13), 2000-2005.

36. Kurbanoglu, E. B.; Zilbeyaz, K.; Taskin, M.; Kurbanoglu, N. I., Total production of (*R*)-3,5-bistrifluoromethylphenylethanol by asymmetric reduction of 3,5-bis (trifluoromethyl)acetophenone in the submerged culture of *Penicillium expansum* isolate. *Tetrahedron: Asymmetry* **2009,** *20* (23), 2759-2763.

37. Wang, P.; Cai, J.-B.; Ouyang, Q.; He, J.-Y.; Su, H.-Z., Asymmetric biocatalytic reduction of 3,5-bis(trifluoromethyl)acetophenone to (1*R*)-[3,5-bis(trifluoromethyl)phenyl]ethanol using whole cells of newly isolated *Leifsonia xyli* HS0904. *Applied microbiology and biotechnology* **2011,** *90* (6), 1897-1904.

38. Wang, N.-Q.; Sun, J.; Huang, J.; Wang, P., Cloning, expression, and directed evolution of carbonyl reductase from *Leifsonia xyli* HS0904 with enhanced catalytic efficiency. *Applied microbiology and biotechnology* **2014,** *98* (20), 8591-8601.

39. Ouyang, Q.; Wang, P.; Huang, J.; Cai, J.; He, J., Efficient enantioselective synthesis of (*R*)-[3,5-bis(trifluoromethyl)phenyl]ethanol by *Leifsonia xyli* CCTCC M 2010241 using isopropanol as co-substrate. *Journal of microbiology and biotechnology* **2013,** *23* (3), 343-350.

40. Wang, N.; Huang, J.; Luo, H.; Wang, P.; Li, J., Purification and characterization of a new carbonyl reductase from *Leifsonia xyli* HS0904 involved in stereoselective reduction of 3,5-bis(trifluoromethyl)acetophenone. *Journal of Molecular Catalysis B: Enzymatic* **2013,** *92*, 1-6.

41. Huisman, G. W.; Liang, J.; Krebber, A., Practical chiral alcohol manufacture using ketoreductases. *Current opinion in chemical biology* **2010,** *14* (2), 122-129.

42. Liang, J.; Mundorff, E.; Voladri, R.; Jenne, S.; Gilson, L.; Conway, A.; Krebber, A.; Wong, J.; Huisman, G.; Truesdell, S., Highly enantioselective reduction of a small heterocyclic ketone: biocatalytic reduction of tetrahydrothiophene-3-one to the corresponding (*R*)-alcohol. *Organic Process Research & Development* **2010,** *14* (1), 188-192.

43. Woodley, J. M., New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends in biotechnology* **2008,** *26* (6), 321-327.

CHAPTER 2

TERNARY COMPLEX CRYSTAL STRUCTURE OF SECONDARY ALCOHOL DEHYDROGENASES FROM THE *THERMOANAEROBACTER ETHANOLICUS* MUTANTS C295 AND I86A PROVIDES BETTER UNDERSTANDING OF CATALYTIC MECHANISM[1]

[1] Dinh, T. T.; Rahn, K. T.; Phillips, R. S. To be submitted to *ACS Catalysis*

Abstract

Thermoanaerobacter ethanolicus alcohol dehydrogenase (TeADH) can catalyze the reversible reduction of ketones into corresponding secondary alcohols using NADPH as cofactor. The enzyme contains a zinc ion which is tetrahedrally coordinated by four amino acids, one of which is Glu-60. Hydride transfer is proposed to occur between substrate and cofactor in the ternary complex of substrate-enzyme-cofactor. However, there have been no ternary complex of TeSADH crystal reported. While zinc is essential for enzyme catalytic action, its coordination numbers and geometries during catalysis are still up to debate. Previous research by X-ray absorption spectroscopy (XAS) implied the existence of pentacoordinate zinc intermediates and the involvement of Glu-60 in zinc coordination as well as catalytic mechanism. Here, we report a new condition to crystallize TeSADH-I86A mutant complexed with cofactor as well as either 2-pentanol or 3-methylcyclohexanol. Another mutant, C295A, was crystallized by previously published conditions, along with DMSO, a transition-state analogue inhibitor. All substrates, inhibitor and cofactor were incorporated by soaking method and, to the best of our knowledge, this was the first time structures of TeSADH ternary complexes were obtained. Our results prove the formation of pentacoordinate zinc before the displacement of water molecule. Moreover, the crystal structures show the dynamics and interactions of cofactor at the active site upon substrate binding, and, more importantly, the presence of a proton relay system which was not demonstrated in previous structures.

Keywords: *Thermoanaerobacter ethanolicus*, secondary alcohol dehydrogenases, X-Ray crystallography, zinc coordination, pentacoordinate zinc, catalysis, proton relay system, ligand exchange, reaction mechanism

1. Introduction

The minority of enzymes can be active in harsh conditions of pH, temperature, pressure, etc. Proteins that are stable at high temperature can be found in thermophilic bacteria such as *Thermoanaerobacter ethanolicus*. This species was first discovered and isolated in Yellow Stone Park hot spring by Lee et al.[1] and named *Clostridium thermohydrosulfuricum* at first[2]. It was afterwards renamed as *Thermoanaerobacter. ethanolicus* strain, 39E[3]. Its genome sequence is available under the name *T. ethanolicus* 39E at http://genome.ornl.gov/microbial/teth_39e/. Proteins of this thermophilic species attracts some attention as industrial need of enzymes which can work at high temperature, especially Secondary Alcohol Dehydrogenase (SADH). This enzyme is a member of medium-chain dehydrogenase/reductases (MDRs) family. MDRs while having high sequence identity and homology, at 75% and 88%, respectively, are different in thermal stability[4]. Similar to other MDRs, each chain of *T. ethanolicus* Secondary Alcohol Dehydrogenase (TeSADH) contains 352 amino acid residues. It only shares 27 % sequence identity with mammalian horse liver ADH (HLADH)[5].

Many alcohol dehydrogenases (ADHs), including MDRs, can reduce a prochiral carbonyl compound into one enantio-specific corresponding alcohol. Owning to this stereo-selective reduction, MDRs are promising catalysts for asymmetric syntheses. The current model proposes that there are two catalytic pockets at the active site, namely small and large pockets. The two substituents of the prochiral ketone will bind to these two pockets according to their sizes[6]. If the bound cofactor positions on the *re* face of the ketone, its hydride will attack from this side to produce the (*S*)-alcohol, assuming that the large substituent has higher priority than the smaller counterpart, according to the Cahn-Ingold-Prelog rule. TeSADH not only has high thermostability and tolerance toward

solvent but also wide substrate specificity. It has been used to synthesize some significant chiral substances[6-8].

Because of its potential applications in chiral syntheses, there has been studies to broaden the substrate specificity of TeSADH[9]. The wild-type showed almost no activity toward acetophenone and 1-(*R*)-phenylethanol derivatives[10]. This is due to the fact that the small pocket is much more hydrophobic than the larger one. As a result, the phenyl ring preferentially binds to the former though this pocket in the wild-type is inherently not big enough to accommodate aromatic rings. Musa et al. developed TeSADH mutants by replacing several residues to enlarge the small pocket[11]. Among those, the I86A mutant while remaining active toward aliphatic alcohols, showed activity in the presence of acetophenone and 1-(*R*)-phenylethanol, but not 1-(*S*)-phenylethanol. Compared to wild-type enzyme, TeSADH-I86A allows the binding of more sterically demanding substrates like benzylic and heteroaryl ketones and converts them into the corresponding (*R*)-alcohols[11].

Another promising use of TeSADH is its ability to reduce ethynyl alkyl ketones as its products, α-ethynyl alcohols, are important intermediates in organic syntheses. The problem was that the enzyme being irreversibly inactivated after the reaction. Single mutant C295A was created since Cys-295 was thought to form covalent bond with the terminal alkyne via Michael addition, leading to inactivation[12]. While the mutant is still inactivated by ethynyl alkyl ketones, it gave better results than that of wild-type for several substrates, even being active toward substrates that did not react with the mutant. Cys-295 along with Ile-86 and Ala-85, which are part of the small pocket are highly but not absolutely conserved among ADHs since they do not participate directly into catalytic

action[12]. The roles of these residues in substrate binding have not been clear. Nevertheless, the changing from Cys to Ala at position 295 opens the small pocket, enabling the binding of larger alkyl group. Furthermore, this favors α-branching, for example while C295A is active in the presence of ethynyl *sec*-pentyl ketone while the wild-type does not.

Crystal structure of the wild-type TeSADH was first reported by Korkhin[4, 13], in space group $P6_5$ and complex with $NADP^+$ (nicotinamide-adenine dinucleotide phosphate) at 2.5 Å resolution (pdb code: 1YKF). Li et al[14] used a different condition that resulted in crystals of different space group $P2_12_12_1$ and complex with (*S*)-2-butanol (pdb code: 1BXZ) at 3 Å but missing cofactor. Recently, Vidal et al. released an apo-enzyme structure (pdb code: 6SDM) at 2.85 Å resolution. For mutants, the I86A structure was solved by Protsko et al.[15], which lacks substrate due to weak electron density and was not uploaded on http://www.rcsb.org/. To the best of our knowledge, until this point there has been no ternary complex of TeSADH with both a substrate and cofactor (NADP) published. Here, we report crystallization by new conditions of TeSADH-I86A and complexes with 2-pentanol and 3-methylcyclohexanol, as well as TeSADH-C295A with dimethyl sulfoxide (DMSO) by a previously published condition. All crystals were soaked with cryoprotectant solutions containing NADP to incorporate this cofactor. These crystal structures help to elucidate substrate-enzyme-cofactor interactions at its active site and demonstrate the mechanism of ligand exchange on the catalytic zinc.

2. Materials and methods

2.1. Materials

Kanamycin, ampicillin, and NADPH tetrasodium salt were purchased from Sigma-Aldrich. Red agarose column matrices were purchased from GE Healthcare Life Sciences

(USA). All other reagents used in this study were obtained from general commercial sources.

2.2. Strains, plasmids, and culture conditions

The plasmids of TeSADH-C295A and TeSADH-I86A mutants were introduced into *E. coli* BL21(DH5α) by heat shock before the transformed cells were streaked on LB agar plates supplemented with 50 μg/ml kanamycin 100 μg/ml ampicillin and incubated at 37 ^{o}C for 12 h.

2.3. Preparation of cell-free extracts

A single colony for each mutant was cultivated overnight at 200 rpm and 37 ^{o}C in 8 mL LB media. Subsequently, 7 mL of this starter culture was added to 1 L of LB media containing 50 mg kanamycin and 100 mg ampicillin in a 2 L Erlenmeyer flask, and incubated at 200 rpm and 37 ^{o}C overnight. The cells were harvested by centrifugation at 4000 rpm for 15 mins, resuspended in buffer A (Tris-HCl buffer 50 mM, pH 8.0, 10 μM $ZnCl_2$, 5 mM DTT) and then lysed by ultrasonication (100 W, pulse 1 min) in 4 one-minute intervals (with 10 mins rest in between). The cells debris were removed via centrifugation at 4,000 rpm for 90 mins, to obtain the cell-free extract.

2.4. Protein purification

Subsequently, proteins were purified using a Low Pressure (LP) chromatography system (Bio-Rad, USA). The Red Agarose (GE Healthcare, USA) column was first recharged with 50 mL solutions in the following order: DI water, 4M urea, DI water, 4M NaCl, DI water at a flow rate of 1.0 mL/min before being equilibrated with buffer A at the same flow rate. After the cell-free extract was then loaded onto the column, buffer A was used to wash it until the absorbances returned to a flat baseline. The protein was eluted with buffer A plus 0.2 M $NaClO_4$ and was collected with a fractional collector at every 1 mL. Peak fractions

were pooled together (from 8-12 mL fractions) and the protein concentration were adjusted to approximately 10 mg/mL for crystallization.

2.5. Protein crystallization

2.5.1. TeSADH-C295A

The enzyme was crystallized by vapor diffusion in hanging drops by mixing 1 µl of the protein solution (10 mg/ml) with 1 µl of 50 mM NaCl, 50 mM Tris-HCl (pH = 8.3), 14% PEG 4000, and equilibrated at 295 K over 1 mL of this solution. Crystals, which fully developed after 3-4 weeks (Supplemental data Figure S2.1), were flash-cooled in liquid nitrogen with crystallization solution plus 20% EDG mixture (1:1:1 of ethylene glycol, DMSO, and glycerol) as cryoprotectant before data collection. NADP was also added to a final concentration of 3 mM. The crystals belong to the tetragonal space group P 2_1 2_1 2_1 with unit cell dimensions of a = 79.59 Å, b = 125.13 Å, c = 137.14 Å, $\alpha = \beta = \gamma = 90°$.

2.5.2. TeSADH-I86A

Screening. Several experimental conditions published previously[4, 14-15] to crystallize TeSADH-I86A were tried. However, none of these resulted in good quality crystal. Therefore, high-throughput screening was performed using Qiagen NeXtal Classics I, Classics II, PEGs I and PEGs II suites. The screen was conducted using a sitting-drop vapor diffusion method with a 10 mg/mL protein solution on four 96-well 2-drop MRC crystallization plates. As a result, the selected condition for optimization was 0.4 M KCl, 0.1 M HEPES pH 7.5, 15% (w/v) PEG 3350.

Optimization. The optimal condition to crystallize TeSADH-I86A is 0.6 M KCl, 12% PEG 3350, 50 mM HEPES buffer pH 7.5.

2.6. Data collection

Crystals, which fully developed after 3-4 weeks, were flash-cooled in liquid nitrogen with crystallization solution plus 20% mixture of ethylene glycol, glycerol, and dimethylsulfoxide (DMSO) (1:1:1). For TeSADH-I86A, the substrate, either 2-pentanol or 3-methyl cyclohexanol, was added to cryoprotectant solution before crystal soaking. NADP were also added to a final concentration of 3 mM. Diffraction data were collected at 100 K at the 22-BM beamline of the SER-CAT at the Advanced Photon Source, Argonne National Laboratory. The wavelength was at 1 Å. One crystal (0.1 × 0.1 × 0.1 mm) was used to collect data at 100 K. All data sets were collected with 1 s exposure/0.25°/ frame (rotation of ω) using 249.14 mm crystal to detector distance. The data were integrated and reduced using XDS[16] and scaled with the program XSCALE[17]. Ellipsoidal truncation was carried out on http://staraniso.globalphasing.org.

2.7. Structure determination and refinement

To determine the TeSADH-C295A structure, Molecular replacement (Phaser-MR-simple one-component interface)[18] by PHENIX was employed. The search model was the crystal structure of NADP-dependent alcohol dehydrogenase[13] from *Thermoanaerobium brockii* (pdb code: 1ykf). The crystals of TeSADH-C295A and I86A belong to orthorhombic space group of $P2_12_12_1$, with unit cell parameters roughly a = 79.59 Å, b = 125.13 Å, c = 137.15 Å, α = β = γ = 90°. The Matthew coefficient of the 3 structures were 2.16 ($Å^3$/Da), with solvent content of about 0.43, proposing the presence of a tetramer in the asymmetric units. After that, PHENIX.autobuild[19] was performed to build the structure and initial experimental map. The model was then refined with PHENIX.refine[20] until R_{free} reached a minimum. For the first refinement cycle, rigid body refinement was used to improve the positional and orientational parameters of the four subunits as well as strict

non-crystallographic symmetry (NCS). The map was of good quality and used for further model building in COOT[21] manually. The electron density map calculated at 1 sigma was well connected. Water, ions, and ligands were added with COOT[21].

Structure validation and deposition. The quality of the structure was evaluated with the validation tools included in the programs COOT[21] and MOLPROBITY[22]. All secondary structure work and figures in this paper was prepared by PyMOL (The PyMOL Molecular Graphics System, Version 1.84, Schrödinger, LLC.). Atomic coordinates of TeSADH-C295A at 1.86 Å with DMSO and NADP have been deposited in the PDB and are accessible under PDB ID 7JNQ.pdb. TeSADH-I86A complexes with 2-pentanol and 3-methylcyclohexanol at 1.98 Å and 2.34 Å were deposited under PDB ID 7JNU and 7JNS, respectively.

Create anomalous maps for zinc. The same data sets were integrated by XDS but this time with Friedel's rule set to false to get the anomalous signals. The scaling was done without merging and the data were then subjected to MR-SAD by Phaser-EP in phenix[23-24] with 3.95 Å resolution cutoff. The crystal structures were utilized as initial model for phasing. After that calculating unusual map coefficient (cmapcoeff) was conducted by in ccp4i2 package[25]. The resulting map was then overlaid with the model to verify the presence of Zn^{2+}. Three Zn^{2+} out of four chains (chain B, C, and D) were found even at very high contour $\sigma = 7$ (Supplemental data, Figure S2.2).

3. Results and Discussion

3.1. Crystal and refinement statistics

All crystal structures were refined to satisfactory R_{free} and stereochemistry parameters. X-Ray diffraction data of TeSADH-C295A+DMSO, TeSADH-I86A+2-pentanol, and

TeSADH-I86A+3-methylcyclohexanol are summarized in Table 2.1. All were in presence of NADP.

Table 2.1. Summary of crystal and refinement statistics of TeSADH-C295A+DMSO, TeSADH-I86A+2-pentanol, and TeSADH-I86A+3-methylcyclohexanol complexes (statistics for the highest-resolution shell are shown in parentheses)

	Crystal data		
	C295A+DMSO (7JNQ)	I86A+2-pentanol (7JNU)	I86A+3-methylcyclohexanol (7JNS)
Unit cell dimension	a = 79.59 Å b = 125.13 Å c = 137.14 Å	a = 80.02 Å b = 124.57 Å c = 166.98 Å	a = 79.82 Å b = 122.85 Å c = 165.15 Å
	$\alpha = \beta = \gamma = 90°$		
Space group (number)	$P\ 2_1\ 2_1\ 2_1$ (19)		
MW Da (residues)	150 (1449)		
	Collection data		
	7JNQ	7JNU	7JNS
Wavelength (Å)	1	1	1
Resolution range (Å)	68.84 - 1.87 (1.937 - 1.87)	39.48 - 2.4 (2.486 - 2.4)	49.28 - 2.34 (2.424 - 2.34)
Total reflections	1695471 (167123)	966903 (97649)	464993 (48308)
Unique reflections	113585 (925)	65675 (6453)	69165 (350)
Redundancy	14.9 (14.9)	14.7 (15.1)	6.7 (7.1)
Completeness (%)	61.22 (8.24)	99.91 (99.83)	73.41 (5.13)
Mean I/sigma(I)	10.65 (0.40)	17.15 (1.91)	9.40 (0.63)
Wilson B-factor	31.37	51.51	53.67
$R\text{-}_{merge}$	0.2295 (8.264)	0.1372 (1.799)	0.1705 (3.375)
$R\text{-}_{meas}$	0.2376 (8.556)	0.1422 (1.861)	0.185 (3.64)
$R\text{-}_{pim}$	0.06128 (2.205)	0.03691 (0.4743)	0.07091 (1.355)
$CC_{1/2}$	0.999 (0.217)	0.999 (0.665)	0.997 (0.175)
CC*	1 (0.597)	1 (0.894)	0.999 (0.545)
Refinement			
Reflections used in refinement	69534 (925)	65663 (6453)	50827 (350)
Reflections used for $R\text{-}_{free}$	2000 (27)	2000 (197)	2000 (14)

Resolution range (Å)	68.84 - 1.87		
R_{work}	0.2391 (0.3966)	0.2339 (0.3498)	0.1814 (0.2784)
R_{free}	0.2550 (0.469)	0.2503 (0.3804)	0.2303 (0.2825)
CC(work)	0.889 (0.364)	0.945 (0.676)	0.958 (0.762)
CC(free)	0.881 (0.469)	0.949 (0.594)	0.918 (0.418)
No. of			
Non-hydrogen atoms	10984	10763	10956
Water	296	74	171
Rms deviation from ideal geometry			
Bond length (1-2) (Å)	0.004	0.006	0.007
Angle (°)	1.5	0.66	0.90
Dihedral (°)	18.763		
Ramachandran plot statistics (%)			
Residues in most favor regions	96.00	96.28	95.27
Residues in allowed regions	3.79	3.65	4.44
Residues in disallowed regions	0.21	0.07	0.29
Rotamer outliers (%)	0.09	0.36	0.91
Clashscore	7.97	3.36	13.08
Average B-factor ($Å^2$)	39.07	69.07	53.65

3.2. Overall structure

The final model of TESADH-C295A consists of a homotetramer in the asymmetric unit. The enzyme is about 85 Å wide and long and contains 351 residues for each chain, water molecules, 4 NADP molecules, and 4 zinc ions. It consists of 4 β-sheet, 4 βαβα units, 2 β-hairpins, 9 β-bulges, 17 β-strands, 13 α-helices, 32 β-turns, and 2 γ-turns (Supplemental data Figure S2.3). The loop from residue 218 to 220 is disordered in the crystal, presenting significantly weak electron density in this region.

3.3. Catalytic site

TeSADH-C295A

DMSO is a stable transition-state analogue for TeSADH substrate and was proved to inhibit the enzyme activity in the presence of 2-propanol[26]. For C295A crystal, because the

concentration of DMSO in the cryoprotecting solution was high, at 6.7 % (v/v), in the structure, 33 DMSO molecules were found, four of which occupy the four active sites of the four chains (A, B, C, and D), coordinating with zinc. To ensure that DMSO and NADP molecules at the catalytic sites are not artifact from model bias or noise, a simulated-annealing OMIT map was run on all NADP and DMSO molecules, including the ones not at the active site. The obtained map proves the presences of these ligands (Supplemental data, Figure S2.4).

TeSADH-I86A

Simulated-annealing OMIT maps for 7JNU and 7JNS for 2-pentanol and 3-methylcyclohexanol as well as NADP cofactor confirm the presences of these ligands (Supplemental data, Figure S2.5). Beside coordination with Zn^{2+}, the two substrates are shown within interaction distance with Ser-39 and Asp-150 (Figure 2.2). For 7JNU, even though a racemic mixture of 2-pentanol was used, only (*R*)-2-pentanol was found at the active site (Figure 2.2, left). When residues were demonstrated in spheres (Figure 2.3), it can be clearly seen that the n-propyl substituent is pointing toward the small pocket which forms by residues like, Ala-85, Ala-86, Asp-150, Leu-294, and Cys-295. The methyl group on the hand is found in the large pocket consisting of Ser-39 and Tyr-267. While in term of size, *n*-propyl is larger than methyl, since the small pocket is more hydrophobic than the large counterpart, let alone Ile-86A was replaced by Ala making it even more hydrophobic, *n*-propyl preferentially binds to the small pocket. Similarly for 7JNS, while racemic mixture of all four possible stereoisomers of 3-methylcyclohexnanol was used, only the configuration of (1*R*,3*R*) configuration was found. Interestingly, the substrate is in chair conformation with the hydroxyl group being in axial position (Figure 2.2, right), which

facilitates the ligation of this group to the zinc center. Stereoelectronic effects dictate that reduction of cyclohexanones by hydride transfer must occur from the top face of the carbonyl to give the axial alcohol, and hence the oxidation must also occur with the axial alcohol, according to the Cieplak effect.

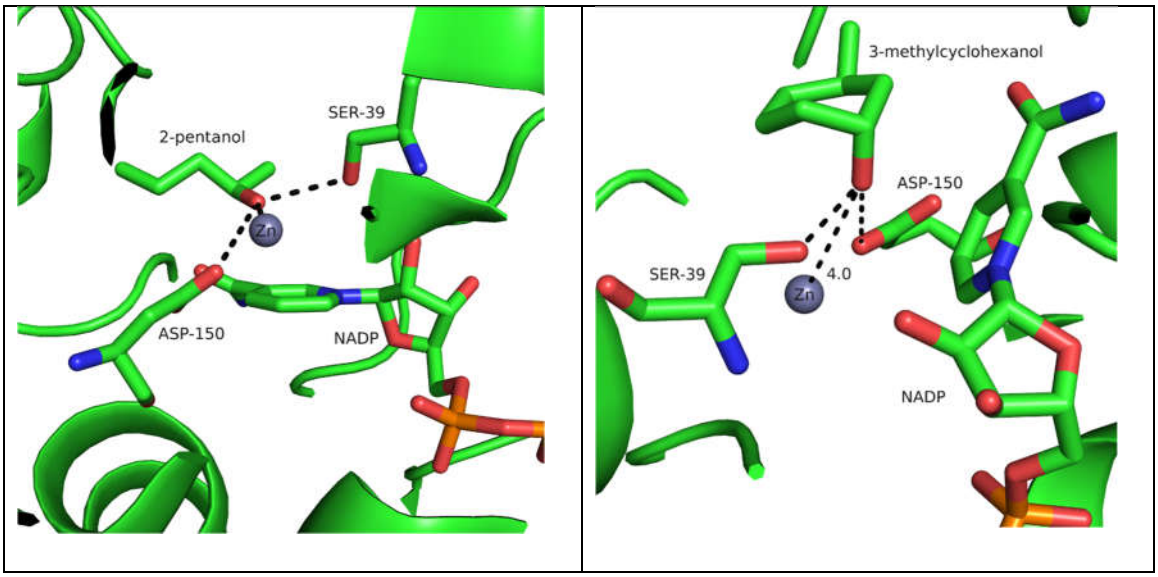


Figure 2.2. Substrate positions and interactions at TeSADH-I86A active site; left: 2-pentanol (7JNU), right: 3-methylcyclohexanol (7JNS)

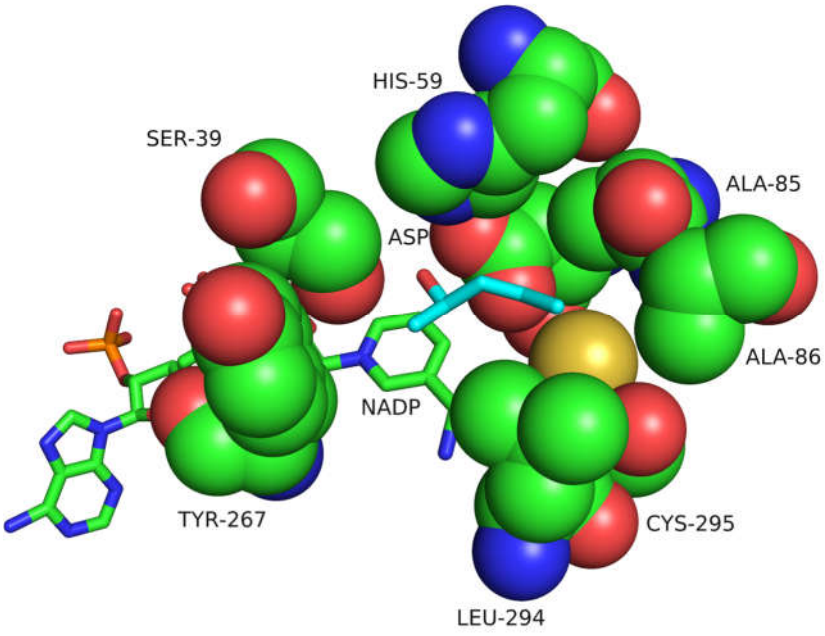
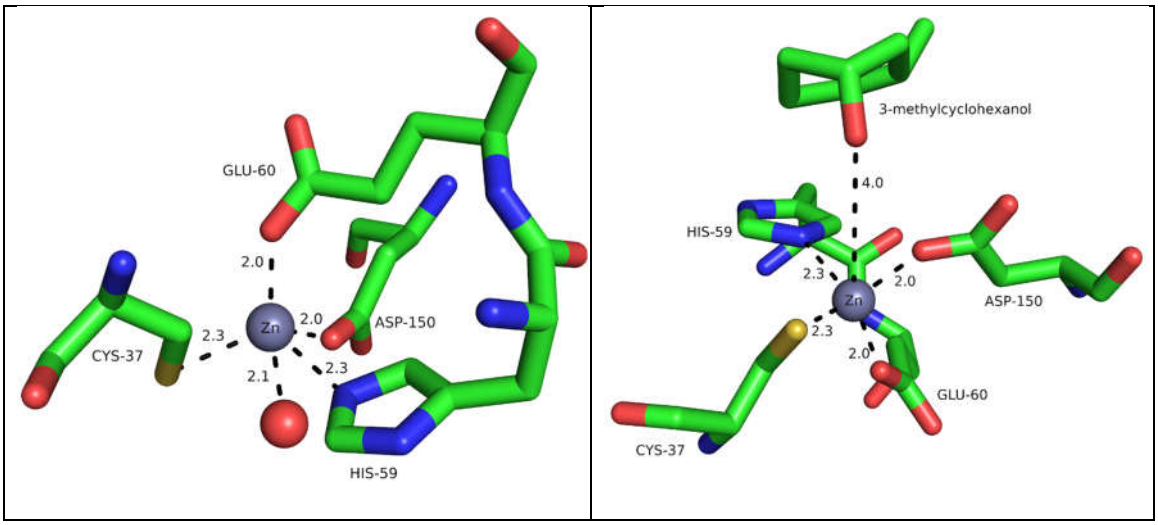


Figure 2.3. Active site of 7JNU

Catalytic Zinc Coordination



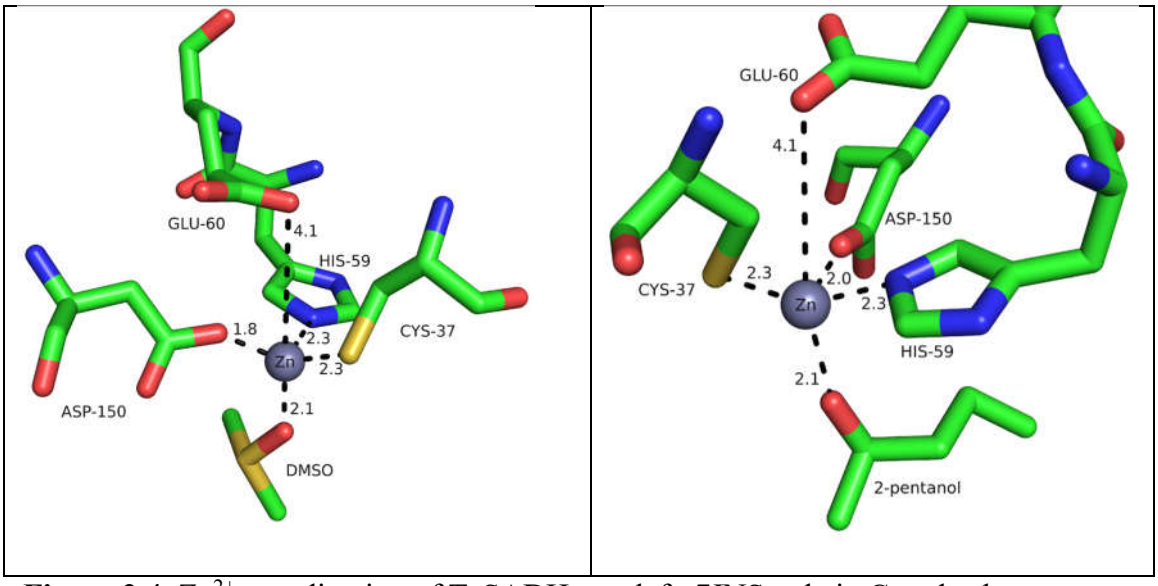


Figure 2.4. Zn^{2+} coordination of TeSADH; top left: 7JNS - chain C, red sphere: water; top right: 7JNS - chain D; bottom left: 7JNQ with coordinated DMSO; bottom right 7JNU with coordinated 2-pentanol.

Interestingly, chain C in 7JNS has water coordinating to zinc instead of the substrates (Figure 2.4, top left). In this subunit, water along with Cys-37, His-59, Glu-60, and Asp-150 arrange to an almost ideal trigonal bipyramidal structure with Glu-60 as axial ligand. In other chains including A, B, and D, while Glu-60 still coordinate with zinc, the substrate 3-methylcyclohexanol is not within coordination distance with the metal ion. In other words, the coordination number of Zn^{2+} before and after substrate binding remain the same. Similar coordination geometry was found in other crystal structures (7JNU and 7JNS). It can be concluded that as soon as either the substrate or analog bind, Glu-60 moves further away from zinc, reducing its coordination number by one. The zinc ion coordination sphere has not change after water is replaced by substrate, hence the latter must displace the former before reactions can take place. All the crystal structures also rule out the hexacoordinated ligands. Last but not least, there was no instance of the substrate and water

being coordinated to the zinc center simultaneously in all crystal structures. This invalidates the suggested model in which water and substrate coordinate with the metal ion simultaneously. Ryde et al.[27] proposed by the theoretical computations that Glu-68 (Glu-60 in TeSADH), one of the most conserved amino acids in alcohol dehydrogenases from various species, might intermittently coordinate to Zn^{2+}, thereby facilitate ligand exchange during catalysis. Moreover, in the structure, Glu-60 forms a weak hydrogen bond with Lys-346 (Figure 2.5, left), enabling the second carboxylate oxygen to coordinate directly to zinc and become the axial ligand in place of a water molecule. When Glu-60 moves away from Zn^{2+}, it comes closed to Lys-340 to make a slightly stronger hydrogen bond (Figure 2.5, right). Kleifeld et al.[28] mutated Glu-60 to Ala which resulted in lower binding affinities of substrate and cofactor for the E60A mutant. This can explain by the loss of ionic interaction between Glu-60 and Lys-346 and as a result, bringing about the increase in position charge in the catalytic site which is required to be neutralized.

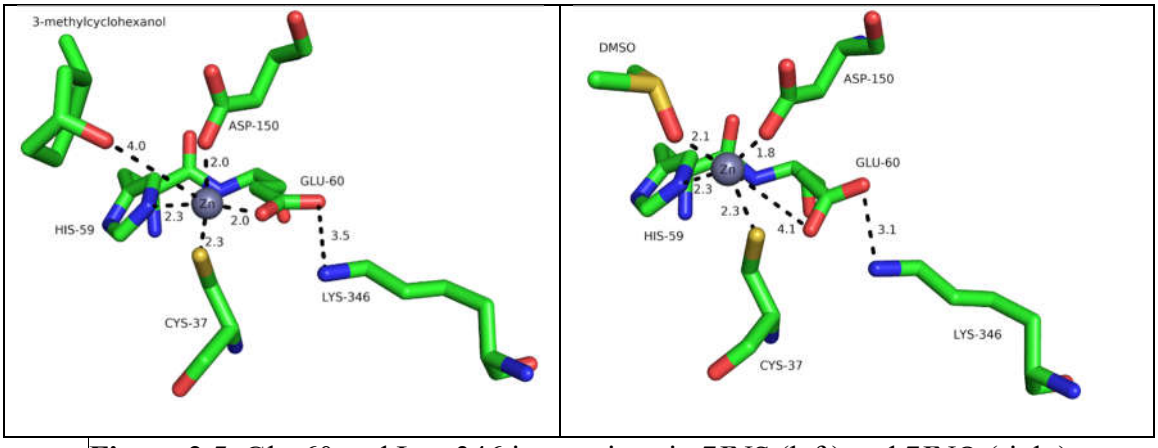


Figure 2.5. Glu-60 and Lys-346 interactions in 7JNS (left) and 7JNQ (right)

Cofactor Interactions, Positioning, and Orientation

As substrate enters the active site, it is sandwiched between Trp-110 residue and the pyridine ring of nicotinamide (Figure 2.6, left). The overlaying of the 1YKF and 7JNU,

showed that the binding of substrate induces the nicotinamide ring rotating by 90º and it places right at the previous position of the primary amide (Figure 2.6, right). After the rotation, the hydride on C4 of NADPH is in position ready to react with the carbonyl if it is in place of the hydroxyl group.

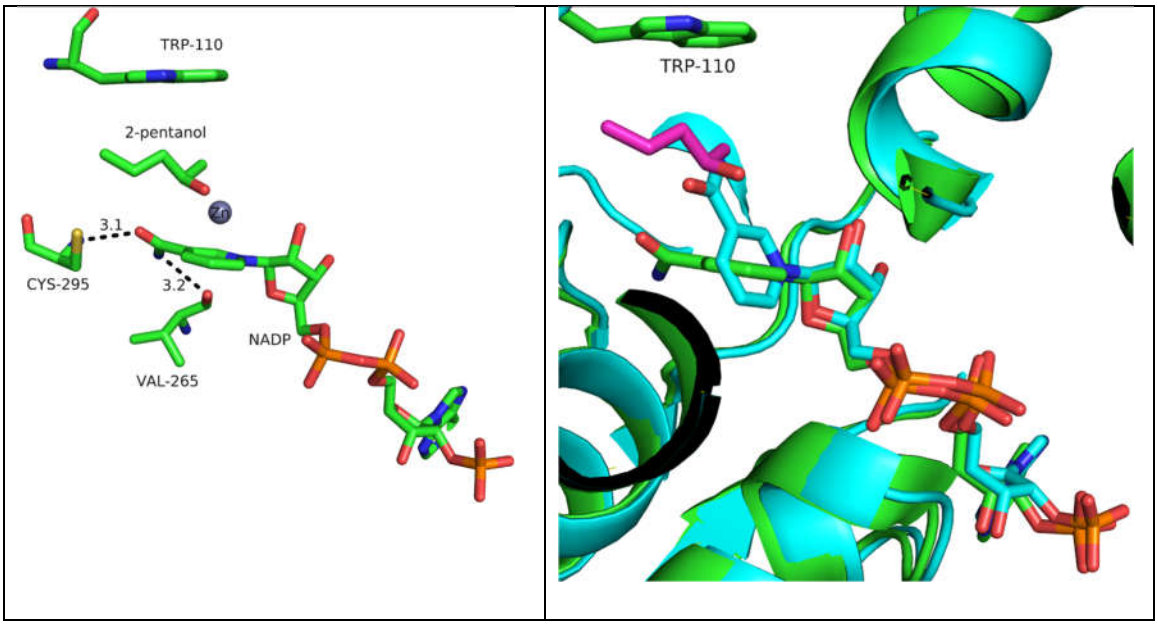


Figure 2.6. Position of 2-pentanol in 7JNU (left) and in overlaying structure of 1YKF and 7JNU with NADP molecules (right)

To investigate the interactions of NADP at the active site, Pymol was used to calculate polar interactions of the cofactor with all residue molecules in the presence (7JNQ, 7JNU, 7JNS) and in the absence of substrates (1YKF). The results revealed that while there were some variations, the four crystal structures have most of the interacting residues in common. More specifically, all structures show interactions from Ser-199 and Arg-200 to the 2''-phosphate group on the adenine ribose, whereas Ile-175 interacts with the 3''-OH of the same ring. In all three but 7JNQ structures, Val-178 shows interaction with disphosphate group while the Asn-266 to Tyr-267 segment forms hydrogen bonds with the 3'-OH of the nicotinamide ribose. All the aforementioned residues help to anchor and

position it into the right orientation (Figure 2.7). However, the most important factor is the orientation of NADPH nicotinamide ring. Upon the binding of substrates, this ring was showed in all three of our crystal structures to rotate and form hydrogen bonds with their side chains, for example Cys-295A or Val-265 in case of 7JNU (Figure 2.6, left). These two interactions were not present in enzyme without substrate (1YKF). For TeSADH-C295A, the mutation at 295 from cysteine to alanine does not affect this position interaction since the hydrogen acceptor is carbonyl instead of the thiol group (Figure 2.7, right).

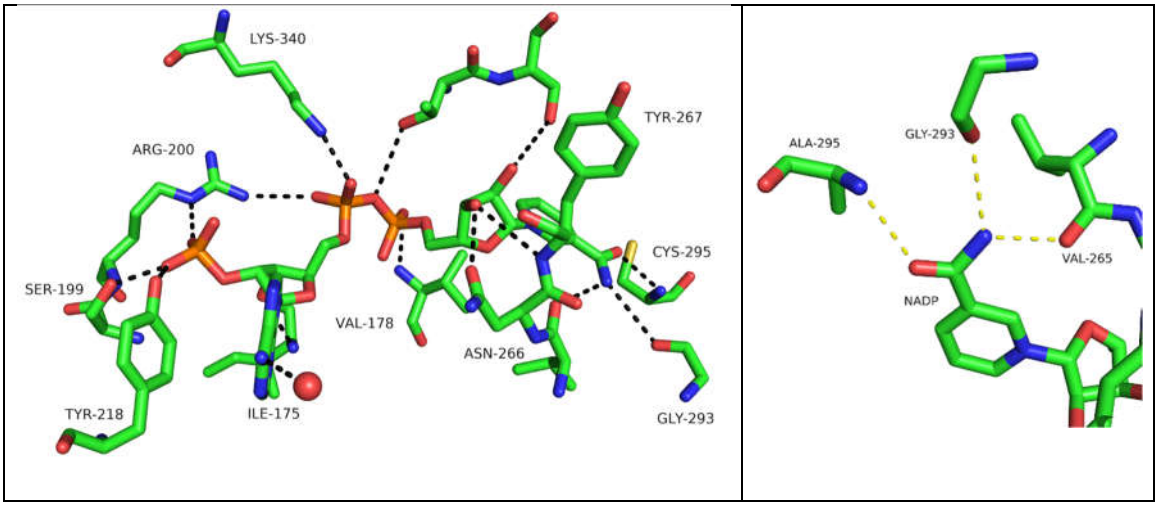


Figure 2.7. NADP interactions with residues in 7JNS (left) and 7JNQ (right)

Proposed Mechanism and the Proton relay system of TeSADH

Previously, the role for Glu-60, which is a highly conserved residue in ADHs, has not been conclusively proven since, to the best of our knowledge, pentacoordinate zinc in a structure of ADHs has not been seen before although it has been suggested based on spectroscopic data[29]. Kleifeld et al. by X-ray absorption spectroscopy (XAS) confirmed the existences of two different pentacoordinate intermediate states, namely TR1 and TR2. However, these results are not conclusive since the EXAFS model is not representative of the active coordination number, instead it only gives the average structure of the whole

sample. In fact, most uncertainty in EXAFS coordination numbers are estimated around ±20-25 % [30]. This is worsened by the fact that zinc ligands were heterogeneous in the structure and enzyme and NADP(H), as well as the lack of data for the steady state phase during which the turnover of the complexes at the active sites can be observed[31]. Moreover, Kleifeld's results cannot rule out the possibility of a sixth ligation by oxygen from a proximal amino acid residue, which is disproven by our crystal structures.

On the basis of our research on X-ray crystallography of TeSADH, we propose the mechanism of action in the oxidation direction as demonstrated in Figure 2.8. Glu-60 reversibly adds to the zinc, which was already coordinated by a water molecule, to form the pentacoordinate structure. When alcohol binds, the water is displaced from the pentacoordinate zinc to give a noncovalent complex with tetrahedral zinc bound by Glu-60. The alcohol then attacks the zinc to form a pentacoordinate zinc, which is unfortunately, not observed. This is likely due to the short lifetime of this intermediate. After that, Glu-60 moves away from the coordination sphere, resulting in a covalent zinc-alkoxide tetrahedral complex. Here, the ligand exchange takes place by the interchange of axial ligands on zinc. However, the issue with our model is that according to the work of Kleifeld et al.[28], TeSADH-E60A mutant only lead to small decrease in the enzyme kinetics and binding affinities of substrate and cofactor relative to the wild type. It was concluded that Glu-60 was displaced by a water molecule during the enzyme action and was not essential for catalysis, yet it is highly conserved in the MDRs family. We hypothesized that in the absence of this residue, the protein may use another amino acid side chain to coordinate with Zn but more experiments must be done to verify this. Nevertheless, numerous studies have demonstrated how dynamic the central zinc and its roles in enzyme catalysis. One

extreme example is the glucose dehydrogenase (GlcDH), another MDR found in *Haloferax mediterranei*[32] in which the zinc atom moves 1.4 Å away from the axial Glu-64 (equivalent to Glu-60 in TeSADH). This type of dramatic movement was though not observed in any of our TeSADH structures.

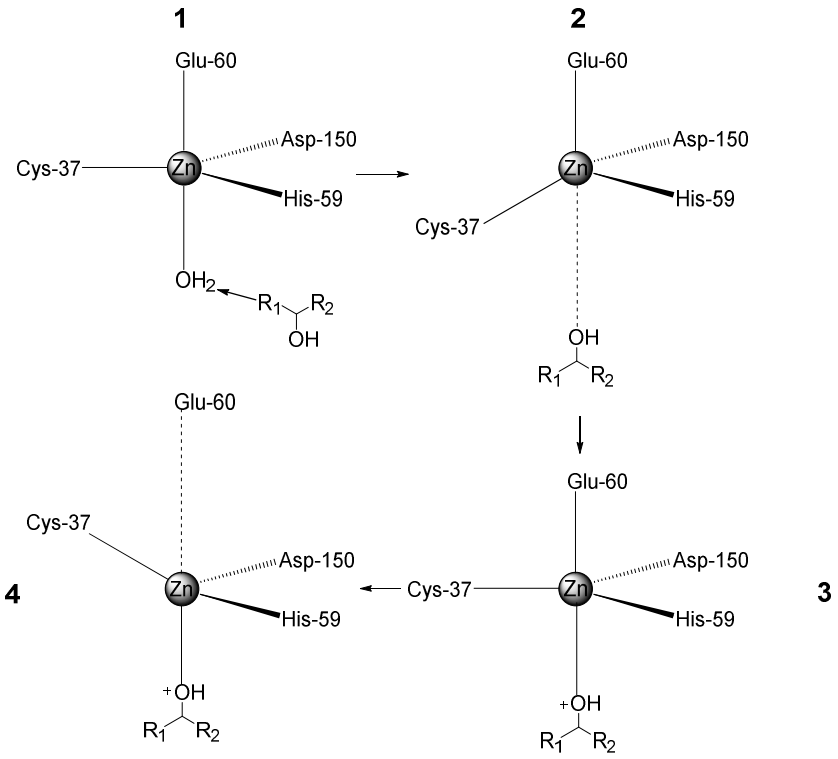


Figure 2.8. Proposed mechanism for the involvement of Glu-60 in the catalytic activity

In the later stage, hydride is transferred to NADP from alcohol, giving rise to zinc bound ketone and neutral NADPH (Figure 2.9). The role of zinc here is stabilizing the transition state by ligation with substrate oxygen and amino acids at the active site. Finally, the ketone was displaced by a water molecule to return the zinc center to original form and NADPH is released, which complete the catalytic cycle. In the oxidation reaction, zinc enhances the deprotonation of secondary alcohol and thereby promoting hydride transfer from the zinc alkoxide intermediate. On the contrary, in the reduction direction, its role is increasing the

electrophilicity of the carbonyl carbon atom. Some members of the ADHs family have been shown to have a proton relay system to transfer proton from alcohol during the oxidation reaction to the surface[13]. For example, HLADH proton shuttle consists of network of hydrogen bonds involving hydroxyl groups from alcohol and Ser-48, the 2'-OH of NADP ribose ring, and His-51 imidazole ring[33-34]. Korkhin et al.[13] did not find this system because in their crystal structures Ser-39 and His-42 are too distant from the 2'-OH of NADP nicotinamide ribose. Our structure, however, exhibited a proton relay system very similar to that of HLADH, which is responsible for accepting proton and shuttling it in and out the active site (Figure 2. 9 & 2.10).

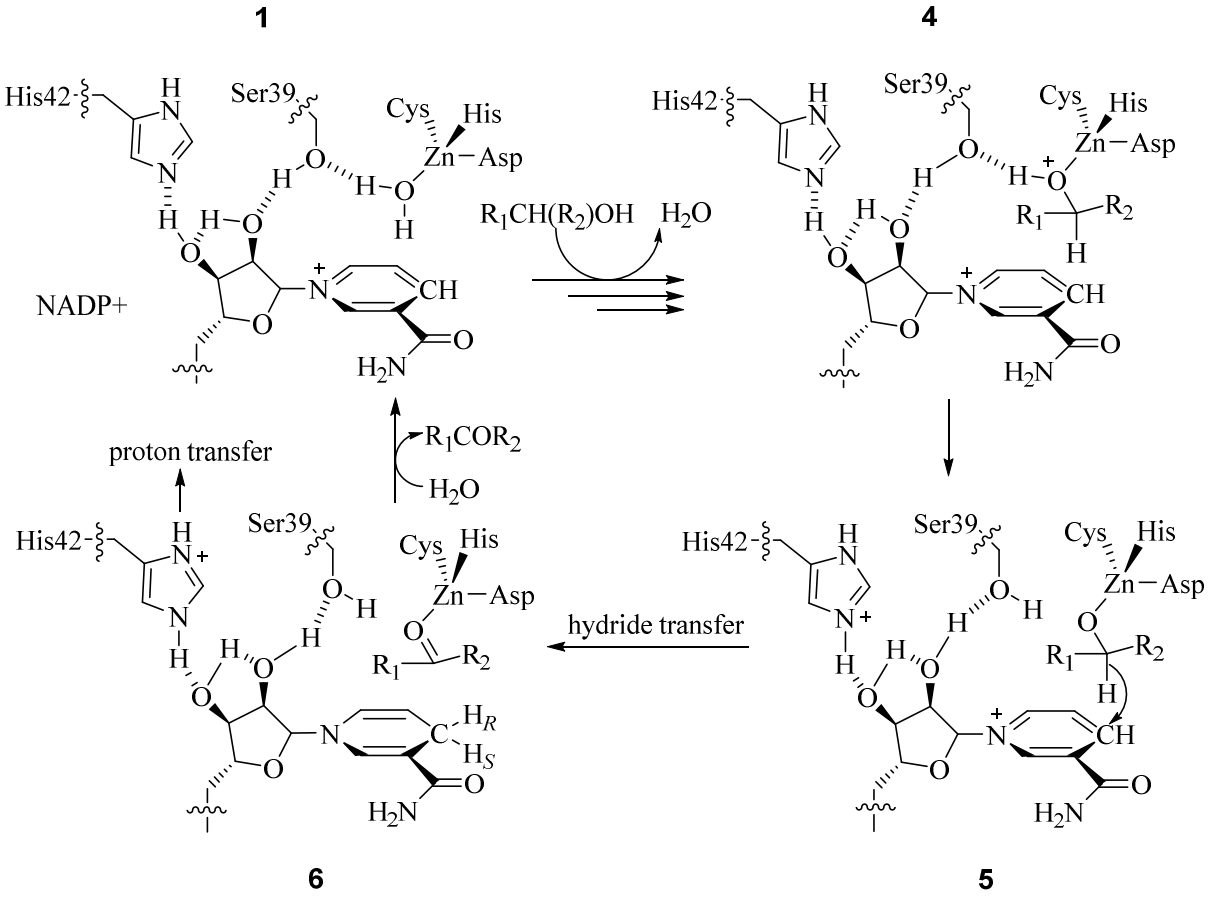


Figure 2.9. Later stage of TeSADH catalytic mechanism coupled with the proton relay system from Ser-39 to His-42

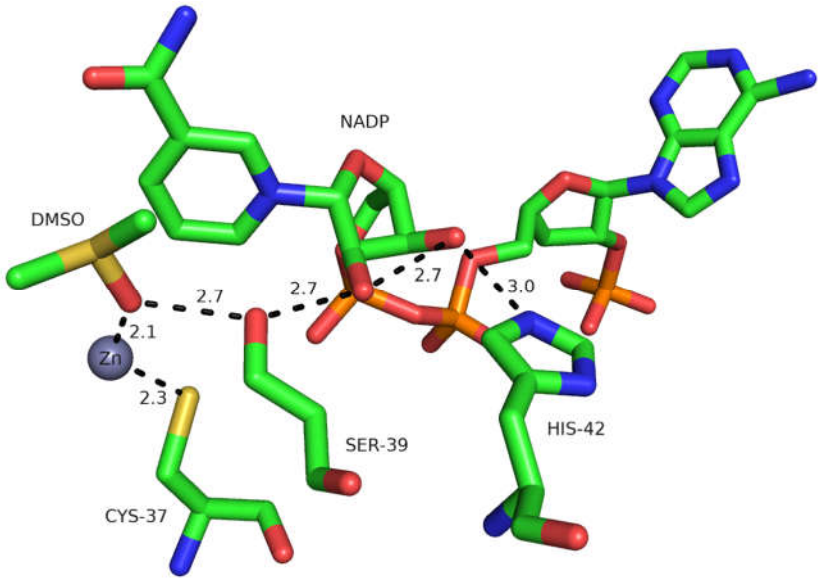


Figure 2.10. Proton relay system in TeSADH

Conclusion

The crystal structure of the ternary complexes of TeSADH, NADP as well as several substrates and inhibitors shed light on the catalytic mechanism of this enzyme. We incorporated DMSO into to enzyme by soaking the crystal with cryoprotecting solution containing this compound. However, the crystal structure of TeSADH-C295A with DMSO did not show a pentacoordinate intermediate but the I86A mutant with 3-methylcyclohexanol exhibit a pentacoordinate Zn with water as one of the five ligands in one of the subunits. Our results invalidate the model in which both alcohol and water molecules coordinate to zinc before Glu-60 is displaced. Furthermore, coordination of catalytic ion center is in trigonal pyramid shape with Cys-37 sulfur, His-59 nitrogen, and Asp-150 oxygen atoms at the corners of the trigonal base while water and Glu-60 oxygen atoms at the apexes. Glu-60 is significant further away from zinc in the absence of substrate (1YKF) suggests that this residue initially is not ligated with zinc. After water coordinated with this cation is displaced by an alcohol, zinc coordination returns to tetrahedral

geometry since the alcohol is still on the second coordination sphere. In the next step, the substrate covalently bonds to Zn^{2+} to form the pentacoordinate complex before Glu-60 moves away, resulting in a tetracoordinated zinc complex. While pentacoordinate zinc has been only demonstrated in other families, for instance carbonic anhydrase[36], our results prove its presence in TeSADH and its necessity for efficient ligand exchange on zinc during catalysis. In term of cofactor interactions, the most important observation was that upon substrate binding, the nicotinamide ring rotate 90º such that it can form hydrogen bonds with and either Cys-295A or Ala-295 and Gly-293 residues. This forces the hydride to position right underneath carbonyl ready to be attacked. Last but not least, the crystal structures also reveal a proton shuttle system from Ser-39 via the 2'-and 3'-OH of NADP ribose to His-42 which were not demonstrated in previous structures. In the future, new techniques are required to prove the pentacoordinate intermediate with substrate binding.

Acknowledgements

Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID and 22-BM beamlines at the Advanced Photon Source, Argonne National Laboratory, and the University of Georgia X-ray diffraction Core Facility (XRDC). SER-CAT is supported by its member institutions (see https://www.ser.aps.anl.gov/www.ser-cat.org/members.html) and equipment grants (S10 RR25528 and S10RR028976) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-3. The XRDC is supported by its UGA member groups

(http://x-ray.uga.edu/) and an equipment grant (S10 OD021762) from the National Institutes of Health.

References

1. Zeikus, J.; Ben-Bassat, A.; Hegge, P., Microbiology of methanogenesis in thermal, volcanic environments. *Journal of bacteriology* **1980,** *143* (1), 432-440.

2. Wiegel, J.; Ljungdahl, L. G., *Thermoanaerobacter ethanolicus* gen. nov., spec. nov., a new, extreme thermophilic, anaerobic bacterium. *Archives of Microbiology* **1981,** *128* (4), 343-348.

3. Lee, Y.-E.; Jain, M. K.; Lee, C.; Zeikus, J. G., Taxonomic distinction of saccharolytic thermophilic anaerobes: description of *Thermoanaerobacterium xylanolyticum* gen. nov., sp. nov., and *Thermoanaerobacterium saccharolyticum* gen. nov., sp. nov.; reclassification of *Thermoanaerobium brockii*, *Clostridium thermosulfurogenes*, and *Clostridium thermohydrosulfuricum* E100-69 as *Thermoanaerobacter brockii* comb. nov., *Thermoanaerobacterium thermosulfurigenes* comb. nov., and *Thermoanaerobacter thermohydrosulfuricus* comb. nov., respectively; and transfer of *Clostridium thermohydrosulfuricum* 39E to *Thermoanaerobacter ethanolicus*. *International Journal of Systematic and Evolutionary Microbiology* **1993,** *43* (1), 41-51.

4. Korkhin, Y.; Frolow, F.; Bogin, O.; Peretz, M.; Burstein, Y., Crystalline alcohol dehydrogenases from the mesophilic bacterium *Clostridium beijerinckii* and the thermophilic bacterium *Thermoanaerobium brockii*: preparation, characterization and molecular symmetry. *Acta Crystallographica Section D: Biological Crystallography* **1996,** *52* (4), 882-886.

5. Eklund, H.; Nordström, B.; Zeppezauer, E.; Söderlund, G.; Ohlsson, I.; Boiwe, T.; Söderberg, B.-O.; Tapia, O.; Brändén, C.-I.; Åkeson, Å., Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *Journal of molecular biology* **1976,** *102* (1), 27-59.

6. Keinan, E.; Hafeli, E. K.; Seth, K. K.; Lamed, R., Thermostable enzymes in organic synthesis. 2. Asymmetric reduction of ketones with alcohol dehydrogenase from *Thermoanaerobium brockii*. *Journal of the American Chemical Society* **1986,** *108* (1), 162-169.

7. Wong, C.; Drueckhammer, D. G.; Sweers, H. M., Enzymatic vs. fermentative synthesis: thermostable glucose dehydrogenase catalyzed regeneration of NAD(P)H for use in enzymatic synthesis. *Journal of the American Chemical Society* **1985,** *107* (13), 4028-4031.

8. Drueckhammer, D. G.; Barbas III, C. F.; Nozaki, K.; Wong, C. H.; Wood, C. Y.; Ciufolini, M. A., Chemoenzymic synthesis of chiral furan derivatives: useful building blocks for optically active structures. *The Journal of Organic Chemistry* **1988,** *53* (8), 1607-1611.

9. Nealon, C. M.; Musa, M. M.; Patel, J. M.; Phillips, R. S., Controlling substrate specificity and stereospecificity of alcohol dehydrogenases. *Acs Catalysis* **2015,** *5* (4), 2100-2114.

10. Ziegelmann-Fjeld, K. I.; Musa, M. M.; Phillips, R. S.; Zeikus, J. G.; Vieille, C., A *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase mutant derivative highly active and stereoselective on phenylacetone and benzylacetone. *Protein Engineering, Design & Selection* **2007,** *20* (2), 47-55.

11. Musa, M. M.; Lott, N.; Laivenieks, M.; Watanabe, L.; Vieille, C.; Phillips, R. S., A single point mutation reverses the enantiopreference of *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase. *ChemCatChem* **2009,** *1* (1), 89-93.

12. Heiss, C.; Laivenieks, M.; Zeikus, J. G.; Phillips, R. S., Mutation of cysteine-295 to alanine in secondary alcohol dehydrogenase from *Thermoanaerobacter ethanolicus* affects the enantioselectivity and substrate specificity of ketone reductions. *Bioorganic & medicinal chemistry* **2001,** *9* (7), 1659-1666.

13. Korkhin, Y.; Kalb, A. J.; Peretz, M.; Bogin, O.; Burstein, Y.; Frolow, F., NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. *Journal of molecular biology* **1998,** *278* (5), 967-981.

14. Li, C.; Heatwole, J.; Soelaiman, S.; Shoham, M., Crystal structure of a thermophilic alcohol dehydrogenase substrate complex suggests determinants of substrate specificity and thermostability. *Proteins: Structure, Function, and Bioinformatics* **1999,** *37* (4), 619-627.

15. Protsko, C.; Vieille, C.; Laivenieks, M.; Prasad, L.; Sanders, D. A.; Delbaere, L. T., Crystallization and preliminary X-ray diffraction analysis of the *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase I86A mutant. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **2010,** *66* (7), 831-833.

16. Kabsch, W., Xds. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 125-132.

17. Kabsch, W., Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 133-144.

18. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (4), 658-674.

19. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L.-W.; Read, R. J.; Adams, P. D., Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography* **2008,** *64* (1), 61-69.

20. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* **2012,** *68* (4), 352-367.

21. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (4), 486-501.

22. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **2018,** *27* (1), 293-315.

23. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 213-221.

24. McCoy, A. J.; Storoni, L. C.; Read, R. J., Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallographica Section D: Biological Crystallography* **2004,** *60* (7), 1220-1228.

25. Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K., CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Structural Biology* **2018,** *74* (2), 68-84.

26. Phillips, R. S., Tailoring the substrate specificity of secondary alcohol dehydrogenase. *Canadian journal of chemistry* **2002,** *80* (6), 680-685.

27. Ryde, U., On the role of Glu-68 in alcohol dehydrogenase. *Protein Science* **1995,** *4* (6), 1124-1132.

28. Kleifeld, O.; Shi, S. P.; Zarivach, R.; Eisenstein, M.; Sagi, I., The conserved Glu-60 residue in *Thermoanaerobacter brockii* alcohol dehydrogenase is not essential for catalysis. *Protein science* **2003,** *12* (3), 468-479.

29. Kleifeld, O.; Frenkel, A.; Martin, J. M.; Sagi, I., Active site electronic structure and dynamics during metalloenzyme catalysis. *Nature structural biology* **2003,** *10* (2), 98-103.

30. Penner-Hahn, J. E., Characterization of "spectroscopically quiet" metals in biology. *Coordination Chemistry Reviews* **2005,** *249* (1-2), 161-177.

31. Plapp, B. V.; Savarimuthu, B. R.; Ferraro, D. J.; Rubach, J. K.; Brown, E. N.; Ramaswamy, S., Horse liver alcohol dehydrogenase: zinc coordination and catalysis. *Biochemistry* **2017,** *56* (28), 3632-3646.

32. Baker, P. J.; Britton, K. L.; Fisher, M.; Esclapez, J.; Pire, C.; Bonete, M. J.; Ferrer, J.; Rice, D. W., Active site dynamics in the zinc-dependent medium chain alcohol

dehydrogenase superfamily. *Proceedings of the National Academy of Sciences* **2009,** *106* (3), 779-784.

33. Andersson P, K. J., Lindström A, Oldén B, Pettersson G, Effect of NADH on the pKa of Zinc-Bound Water in Liver Alcohol Dehydrogenase. *European journal of biochemistry* **1981,** *113* (3), 425-433.

34. Malver, O.; Sebastian, M. J.; Oppenheimer, N. J., Alteration in substrate specificity of horse liver alcohol dehydrogenase by an acyclic nicotinamide analog of NAD^+. *DNA repair* **2014,** *23*, 95-100.

35. Kleifeld, O.; Frenkel, A.; Bogin, O.; Eisenstein, M.; Brumfeld, V.; Burstein, Y.; Sagi, I., Spectroscopic studies of inhibited alcohol dehydrogenase from *Thermoanaerobacter brockii*: proposed structure for the catalytic intermediate state. *Biochemistry* **2000,** *39* (26), 7702-7711.

36. Eriksson, A. E.; Kylsten, P. M.; Jones, T. A.; Liljas, A., Crystallographic studies of inhibitor binding sites in human carbonic anhydrase II: a pentacoordinated binding of the SCN^- ion to the zinc at high pH. *Proteins: Structure, Function, and Bioinformatics* **1988,** *4* (4), 283-293.

CHAPTER 3

THE CRYSTAL STRUCTURE OF THE S154Y MUTANT CARBONYL REDUCTASE FROM *LEIFSONIA XYLI* EXPLAINS ENHANCED ACTIVITY FOR 3,5-BIS(TRIFLUOROMETHYL)ACETOPHENONE REDUCTION[2]

[2] Dinh, T. T.; Li, J.; Phillips, R. S. Submitted to *ChemBioChem*

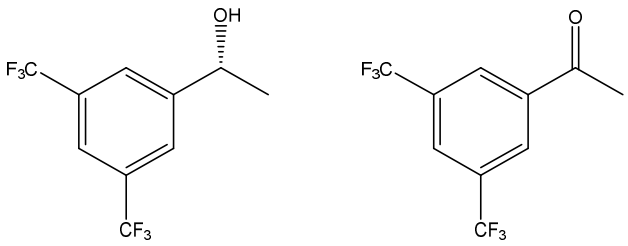
Abstract

Carbonyl reductase from Leifsonia xyli (LXCAR, UniProtKB - T2FLN4) can stereoselectively catalyze the reduction of 3,5-bis(trifluoromethyl)acetophenone (BTAP) to its corresponding alcohol, (R)-[3,5-bis(trifluoromethyl)phenyl]ethanol ((R)-BTPE), which is a valuable chiral intermediate for the synthesis of antiemetic drugs, Aprepitant and Fosaprepitant. Moreover, this protein was found to have a broad spectrum of substrate specificity and can asymmetrically catalyze the reduction of a variety of ketones and keto esters. Even though molecular modelling of this protein was done by Wang et al. (2014), a crystal structure has not yet obtained. In this study, a single mutant, S154Y, which was shown to have higher catalytic activity toward BTAP than that of the wild type, was overexpressed in Escherichia coli BL21 (DE3), purified, and crystallized. The crystal structure of LXCAR-S154Y explains how the mutant enzyme can work with BTAP more efficiently than wild type carbonyl reductase. Furthermore, the structure explains why LXCAR-S154Y can use either NADH or NADPH efficiently as a cofactor, as well as elucidates a proton relay system present in the enzyme.

Introduction

Aprepitant (commercial name: Emend) and Fosaprepitant (commercial name: Ivemend) are popular drugs which have been described for treating side effect of chemotherapy during cancer treatment, including nausea and vomiting.[1-3] Fosaprepitant is merely a pro-drug of Aprepitant in which a phosphate group is added to the latter‘s triazole ring; therefore their syntheses are almost the same [4]. The synthesis pathway requires the key chiral compound, (1*R*)-[3,5-bis(trifluoromethyl)]phenylethanol or (*R*)- BTPE (intermediate 2). While BTPE can be produced by reducing [3,5-bis(trifluoromethyl)]acetophenone or BTAP (Scheme 3.1) with lithium aluminium hydride

($LiAlH_4$) or sodium borohydride ($NaBH_4$), these reagents will lead to a racemic mixture. The following enantiomer separation is painstaking and expensive; hence, it can be said that (*R*)- BTPE is the bottleneck of Aprepitant and Fosaprepitant syntheses. Since the demand for the two drugs has been increasing, the need of efficient and cost-effective production of (*R*)- BTPE has become more and more significant and has been addressed[5-6]. In an attempt to find microorganisms able to selectively reduce BTAP to (R)-BTPE, a project was subsidized to screen soil microbes for this reduction. The screening was detailed by Wang et al.[7] and among the resulting strains, HS0904 worked the best on BTAP which was later identified as *Leifsonia xyli* (or *L. xyli*). Bioinformatic analysis of the *L. xyli* genome revealed that a carbonyl reductase (CAR) gene was responsible for its ability to convert BTAP into (*R*)-BTPE. Analyzing the N-terminal amino acid sequence led to the conclusion that this protein (LXCAR, UniProtKB-T2FLN4) is part of the Short-chain dehydrogenase-reductases (SDRs) family.



(*R*)-BTPE BTAP

Scheme 3.1. Chemical structure of (*R*)-BTPE and BTAP

SDRs are efficient NAD(P)(H)-dependent enzymes for the reduction of a spectrum of ketones to afford enantiopure alcohols.[8-9] Their enantioselectivity has been explained by a model in which there were small and large interacting pockets at the active sites. As

Prelog's rule suggests[10], the smaller pocket binds the smaller substituent from the carbonyl, and vice versa. The relative position of the attacking hydride from the cofactor will decide the product stereochemistry. For instance, if cofactor is on the *si* face of the carbonyl plane, the nucleophilic attack from its hydride will result in a (*R*)-alcohol. SDR enzymes contain typically 250-300 amino acid residues with low residue identities (15~30 %) between the different members, yet show highly conserved three-dimensional structures.[11-12] All SDRs share a typical dinucleotide binding motif, or Rossmann fold, composed of alternating β-strands and α-helices with a central 7-stranded parallel β-sheet packed between two layers of α-helices, and specific residues for the active site, including its highly conserved triad of Ser, Tyr, and Lys residues.[13-14] Most SDR enzymes are homodimers or homotetramers, and all have highly conserved nicotinamide cofactor binding sites, which are formed by a glycine-rich motif on the N-cap of α-helix αA with the sequence motif TGx3GxG in Classical SDRs and TGx2Gx2G in Extended SDRs.[15] The specificity for NADP(H) or NAD(H) is the result of the electrostatics in the cofactor binding site. NADP(H)-dependent SDRs have a conserved arginine or lysine on the C-cap of the second β-strand (β2), while NAD(H)-dependent SDRs are proposed to show negatively charged residues instead.[16] The majority of classical SDRs have been described to favor NADPH, while most extended SDRs prefer NAD(H).[17]

Recently, an S154Y mutant of LXCAR, UniProtKB - T2FLN4) was found to have much higher affinity and activity toward BTAP[18], achieving 99.4% of ee and a yield of 82.5 % for (*R*)-BTPE within 12 h at 1,000 mM (256 g/l) BTAP. A very attracting characteristic of this protein is that it can work with either NAD(H) or NADP(H) as a cofactor with similar K_m and $k_{cat,}$ while many alcohol dehydrogenases (ADHs) are dependent on the more

expensive NADP(H) as cofactor. The fact that LXCAR can react with NAD(H) makes it even more promising as a biocatalyst in cost-effective manufacturing. Wang et al. investigated in the wild type substrate specificity by measuring its activities in the presence of various substrates including ketones, ketoesters, and aldehydes, and then comparing with that of BTAP which was considered as standard (100%).[19] This demonstrated that, while possessing a relatively broad specificity, the enzyme shows very weak activity toward aliphatic aldehyde and alkyl ketones such as acetone, 2-butanone, and 4-methyl-2-pentanone. However, the most striking result was that LXCAR was very active toward trifluoromethylacetophenone (131%) but it has little to no activity toward acetophenone and a number of its mono- or disubstituted halogenated derivatives including 2'-chloroacetophenone, 4'- fluoroacetophenone, and 2',6'-dichloro-3'-fluoroacetophenone. We report here the high-resolution crystal structure of LXCAR-S154Y. The structure explains the high activity of the mutant enzyme for BTAP, and the substrate specificity for NADH and NADPH.

Results and Discussion

Reductase activity

The specific activities of the wild type and mutant are shown in Table 3.1 (top). LXCAR-S154Y has around twice as much activity as that of the wild type, at 23.4 U/mg. The kinetic parameters including K_m, V_{max}, k_{cat}, and k_{cat}/K_m toward BTAP and NADH are summarized in Table 3.1 (bottom). The decrease in K_m along with the increase in k_{cat} values prove that LXCAR-S154Y has higher affinity toward BTAP. Moreover, the mutant showed approximately 12-fold and 4-fold increase in k_{cat}/K_m values toward BTAP and NADH, respectively.

Table 3.1. Kinetic parameters of the WT-LXCAR and LXCAR-154Y toward BTAP (top) and NADH (bottom)

Ligand	BTAP			
	K_m (mM)	V_{max} (U/mg)	k_{cat}[b] (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
WT-LXCAR	0.573±0.008	10.9±0.5	21.1	36.8
LXCAR-S154Y	0.201±0.006	24.7±1.2	114	463

Ligand	NADH			
	K_m (mM)	V_{max} (U/mg)	k_{cat}[b] (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
WT-LXCAR	0.025±0.007	12.2±0.7	23.6	944
LXCAR-S154Y	0.029±0.011	26.6.9±0.9	122	4,207

[b] $k_{cat} = V_{max}/[E]$, where $[E]$ is the molar concentration of the enzyme

Crystal and refinement statistics

The crystal structure was refined to satisfactory R_{free} and stereochemistry parameters. X-Ray diffraction data of LXCAR-S154Y apo enzyme are summarized in Table 3.2. Overall, the electron density map was of high quality and allowed reliable modeling of the monomer. To the best of our knowledge, this is the first structure of LXCAR and among the best resolutions (1.16 Å) in its class. This is particularly significant as SDRs are an enormous superfamily of proteins and these high-resolution structures can be used as models in Molecular Replacement (MR-Phaser) and more importantly homology modeling.

Structural similarity of LXCAR-S154Y to Oxidoreductase family

The MrBump module from CCP4i2 found 5 structural homologues in the Protein Data Bank (PDB)

1. Carbonyl reductase from *Streptomyces coelicolor* (to be published) (PDB code 5h5x), sequence identity = 52%, Z = 31.3
2. 1-(4-Hydroxyphenyl)-Ethanol Dehydrogenase from *Aromatoleum Aromaticum* Ebn1 (PDB code 4urf),[20] sequence identity = 46%, Z = 44.5
3. NADH-dependent ketoreductase ChKRED20 from *Chryseobacterium* sp. CA496 (PDB code 6ixm),[21] sequence identity = 50%, Z = 48.7
4. SDR from *Sinorhizobium meliloti* 1021 (to be published) (PDB code 3tox), sequence identity = 40%, Z = 23.8
5. D-3-hydroxybutyrate dehydrogenase from *Pseudomonas putida* (PDB code 2q2v),[22] sequence identity = 41%, Z = 16.5

As expected, all the structural homologues found are members of the Carbonyl reductase family in the Oxidoreductase superfamily. They have sequence identities, ranging from 40-52%, with LXCAR-S154Y. The high Z-scores indicate that they and the target protein share a very similar secondary structure. The structure of 3tox.pdb was used by Wang et al. for Molecular homology modeling and docking[18]. Among the 5 models, 5h5x with the second highest Z-score and the highest sequence identity was chosen to serve as a template for MR-Phaser in PHENIX.

Table 3.2. Summary of crystal and refinement statistics of LXCAR-S154Y apo enzyme (statistics for the highest-resolution shell are shown in parentheses)

Crystal data	
Unit cell dimension	a = 59.16 Å, b = 59.16 Å, c = 118.92 Å α = β = γ = 90°
Space group (number)	P 4_2 2_1 2 (94)
MW Da (residues)	25049 (251)
Collection data	
Wavelength (Å)	1
Resolution range (Å)	60 - 1.16
Total reflections	868312 (41903)
Unique reflections	73758 (7190)
Redundancy	11.8 (5.8)
Completeness (%)	90.30 (72.35)
Mean I/sigma(I)	13.66 (0.64)
Wilson B-factor	11.08
R_{merge}	0.07011 (2.134)
R_{meas}	0.07324 (2.349)
$CC_{1/2}$	0.999 (0.343)
CC*	1 (0.715)
Refinement	
Reflections used in refinement	66635 (5252)
Reflections used for R_{free}	1810 (143)
Resolution range (Å)	60 - 1.16 (1.130 - 1.160)
R_{work}	0.1644 (0.4021)
R_{free}	0.1768 (0.4173)
No. of	
Non-hydrogen atoms	2133
Water	328
Rms deviation from ideal geometry	
Bond length (1-2) (Å)	0.012
Angle (°)	1.32
Dihedral (°)	14.842
Ramachandran plot statistics (%)	
Residues in most favor regions	96.37
Residues in allowed regions	3.63
Residues in disallowed regions	0.00
Rotamer outliers (%)	0.58
Clashscore	0.28
Average B-factor ($Å^2$)	17.07
Number of TLS groups	3

Overall structure

The final model of LXCAR-S154Y consists of one monomer in the asymmetric unit. The enzyme is of about 50 Å wide and long and contains 251 residues, water molecules, and 3 magnesium ions. It consists of a β-sheet (residues 84-94), 5 βαβα units, 7 β-strands, 14 α-helices, 10 β-turns, and a 2 γ-turn. The loop from residue 192 to 199 is disordered in the crystal, presenting significantly weak electron density. This region is shown in the below section to form hydrogen bonds to the carboxamide from the nicotinamide ring of either NAD^+ or $NADP^+$ to the Ile190 residue. In the early stage of structure refinement of LXCAR-S154Y, the experimental map showed significant positive electron density at the C-terminus, which indicates the presence of metal ions. These coordinates were initially filled with water molecules; however, the map after the following refinement pointed out that was not the case. It was assumed that these sites were Mg^{2+} due to the presence of 0.1 M $MgCl_2$ in the crystallization solution. Wang et al. showed that Mg^{2+} was among several metal ions including Li+, Ca^{2+}, Mn^{2+} and Co^{2+}, that could enhance the enzyme activity. Mg^{2+} was the second best after Mn^{2+}, increasing activity by 1.42-fold.[19]. A calculation using the PISA server (https://www.ebi.ac.uk/pdbe/pisa/) showed that a tetramer, the most stable assembly, is very likely a biologically-relevant oligomeric form. However, it is estimated [19], by gel filtration, a molecular weight of LXCAR about 50 kDa, which on SDS-PAGE showed a clear band at 25 KDa as the monomer subunit. This suggests a facile dimer-tetramer equilibrium exists that may depend on environmental conditions. The activation by divalent cations may reflect an effect of these ions on the dimer- tetramer equilibrium. An extensive interaction between associated subunits of the tetramer buries

approximately 15620 $Å^2$ of the solvent-accessible surface per monomer, which corresponds to about 51% of the total solvent-accessible surface area.

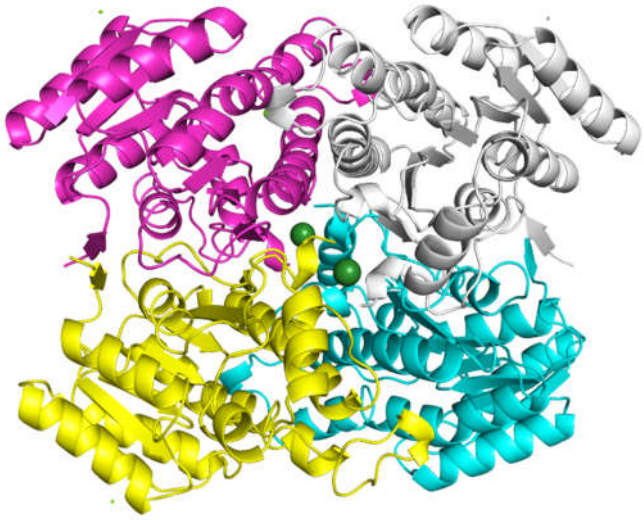


Figure 3.1. Quaternary structure of LXCAR-S154Y. The green spheres are the Mg^{2+} ions (dark green)

Two divalent magnesium cations coordinated at the C-terminal carboxylate groups of the four protomers are relatively rare but not unique for LXCAR[23]. To the best of our knowledge, only two other cases found in ADHs include R-specific alcohol dehydrogenase (RADH) from *Lactobacillus brevis* [24] and dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase from *Salmonella enterica*. [25] In this LXCAR, Mg^{2+} is demonstrated to be part of a dimerization interface. The C-terminal carboxylate of Gln251 and its symmetrical unit counterpart act as mono-dentate axial ligands to the magnesium ions whereas the remaining coordination is provided by oxygen from four water molecules. They altogether form a hexacoordinate octahedral magnesium. Furthermore, in the tetramer modeled by PISA, each of the two magnesiums is ligated by the two chains of LXCAR, which stabilize this quaternary structure (Figure 3.1). Mg^{2+} and water were added manually with COOT[26] and

their presences were confirmed by calculating electron density ($2F_o$-F_c and F_o-F_c) on this cation along with four coordinated water molecules and Gln-251 (Supplemental data, Figure S3.3).

Docking method

Crystals were soaked with a variety of potential ligands, including 1-phenylethanol, 3-trifluoromethyacetophenone, and ethyl 3-hydroxybutyrate. However, while several 1-phenylethanol molecules were found outside the active site of the crystal structure, the one at the active site was not well-defined by electron density. Similarly, several 3-trifluoromethy-acetophenone molecules were found outside of the enzyme active site. Moreover, no density for either NAD^+ or $NADP^+$ cofactor was found. Thus, our crystal structure was superimposed with 5h5x monomer structure by Autodock [27] program, resulting in an RMSD value of 3.261. NAD^+ was copied into the former and then removed and redocked by Autodock vina [28]. The cofactor from 5h5x and from docking position in our crystal structure similarly with an RMSD value of 2.675 (Supplemental data, Figure S3.2). For $NADP^+$ and (*R*)-BTPE, they were docked by Autodock Tools and Autodock vina into the LXCAR crystal structure.

BTPE interactions

All hydrophobic residues which possibly interact with (*R*)-BTPE at its methyl and aromatic ring are investigated. Within 5 Å around BTAP, it can be seen that residue Ile94, Tyr154, Tyr157, Thr192, and Leu194 together form a large pocket accommodating the bistrifluoromethylphenyl group (Figure 3.2, left). Furthermore, residue Leu146, Phe151, and Phe189 produce the small pocket interacting with the methyl group (Figure 3.2, right). If BTAP was present instead of (*R*)-BTPE, the nicotinamide ring of NAD^+ would be on the

si face of this carbonyl, hence the reduction reaction would produce a (*R*)-product. Even though it is likely that NADH will use its pro-*R* hydride to attack a carbonyl, more experiments need to be performed to verify this as described by Pennacchio et al.[29]

The methyl group from docked (R)-BTPE is covered in all directions and has distance, between 3.2 A˚ to 3.7 A˚, to surrounding amino acids including Tyr157, Leu146, Phe189, and Phe151. On the contrary, the phenyl ring from this substrate is not covered completely by hydrophobic side-chains. The most striking observation is the interactions of the trifluoromethyl (CF3) group within the active site. Esterhuysen et al.[30], by computational chemistry, showed that trifluoromethyl, surprisingly, not only can act as a nucleophile but also as an electrophile, generating noncovalent contacts. The polarization resulting from electronic effects can occur if favorable geometry is met. For example, in the model of trifluoromethyl with a phenyl group, one of the fluorine atoms positions beneath the π face and right at the middle of the ring, with the average distance between this atom and six carbons about 3.2 Å. Moreover, two fluorines interact with hydrogen atoms on the ring. All these requirements are met in the interactions of CF3 with Tyr 154 (Figure 3.3). This explains why the mutant enzyme is more active toward BTAP than the wild type. If the Tyr154 is replaced by Ser, all the significant interactions from trifluoromethyl are lost (Figure 3.4, left). Furthermore, the investigation of Esterhuysen et al. on more than 600 protein crystal structures which have trifluoromethylphenyl-containing ligands showed that CF3 group can form hydrogen bonds with NH, CH, OH and SH groups as nucleophile and carbonyl as an electrophile. The most popular motif in enzyme active sites is only one fluorine of CF3 interacting with a carbonyl and its Hα. In LXCAR, these types of interactions can be seen between the trifluoromethyl and Gly95, which help protein to wrap

around the CF3 group and stabilize the conformation. Another interaction from NAD^+ helps to further stabilize (*R*)-BTPE.

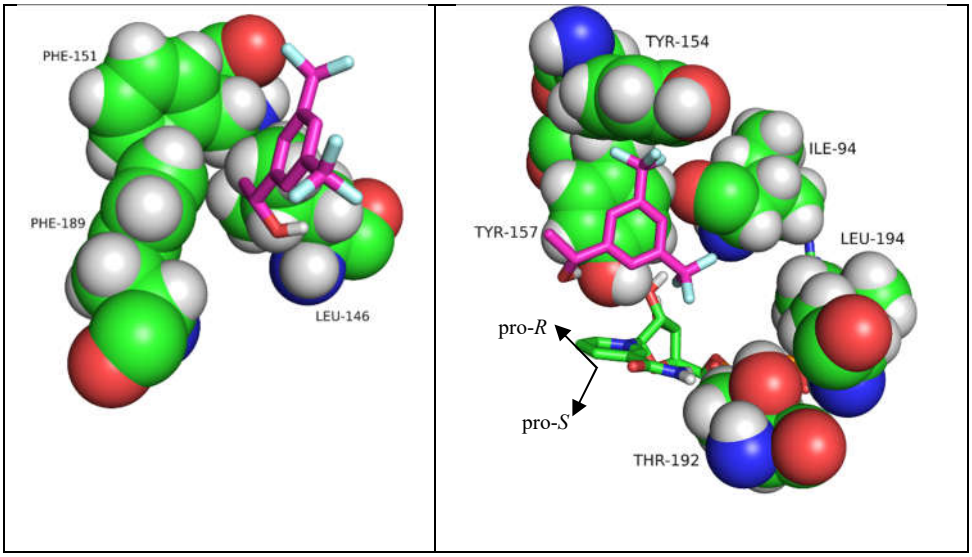


Figure 3.2. (*R*)-BTPE in the large (left) and small pocket (right) surfaces of LXCAR-S154Y

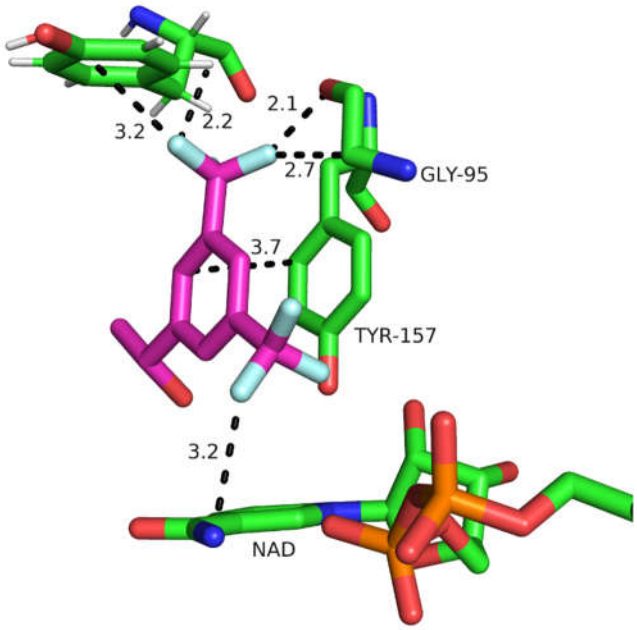


Figure 3.3. Trifluoromethyl (CF_3) interactions in LXCAR-154Y

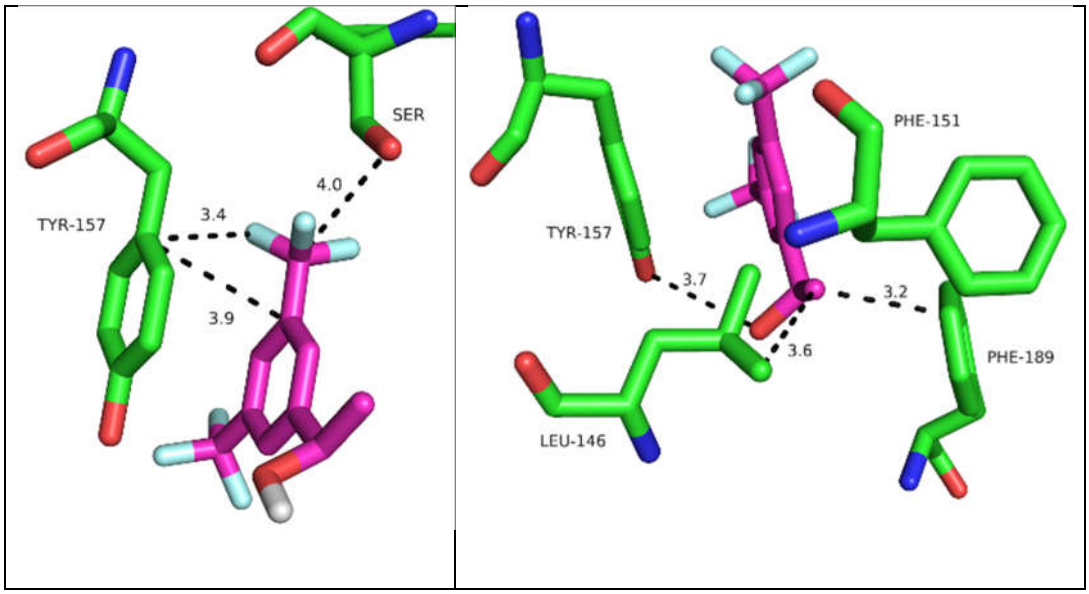


Figure 3.4. left: trifluoromethyl (CF_3) interactions in the wild type; right: interactions of methyl group

Cofactor dependency

SDR proteins can be further categorized into 2 types based on cofactor dependency, namely NAD^+-dependent and $NADP^+$-dependent enzymes. However, LXCAR is unusual in that it uses either NAD^+ or $NADP^+$. The traditional view about the cofactor dependency is that it is determined by 2 residues in the N-terminus region. The first position is the fourth residue at the conserved GXXXGXG motif (Supplemental data, Scheme S3.1) between the strand βA and helix αB (Ser-18 in LXCAR) which decides the preference for $NADP^+$ dependency. The second one is the loop between strand βB and helix αC (Asp-39 and Leu-40 in LXCAR) that determines the NAD^+ dependency.[24] Many SDR enzymes which have two basic arginine residues at both locations use $NADP^+$ as cofactor.[31] In contrast, if, at the second position, an acidic residue like aspartate is present instead of arginine, the SDR enzyme is dependent on NAD^+ as cofactor.[32] Recent studies, however,

have invalidated the requirement of the first location for $NADP^+$ dependency. The wild-type (*R*)-specific alcohol dehydrogenase (RADH) from *Lactobacillus brevis*, while possessing only one arginine at the second location, only uses $NADP^+$ as cofactor[24]. In LXCAR, the fourth residue at the glycine-rich loop (GGGSG from residue 15-21) is Serine instead of Arginine. In the NAD^+-docked structure, Ser18 forms a hydrogen bond with the 2'-OH of adenosine ribose. More importantly, the loop between strand βB and helix αC, including Leu40 and Asp39, interacts with the adenosine moiety of this cofactor (Figure 3.5, top). While the former forms a hydrogen bond with the adenine ring, the aspartate residue forms a hydrogen bond with the 3'-OH of the NAD^+ adenosine ribose. As a phosphate moiety is added to form the $NADP^+$ molecule (Figure 3.5, bottom), it can be seen that Asp39, while still interacting with the 3'-OH of the adenine ribose via a hydrogen bond, can no longer form the same bond with the 2'-OH because of the absence of a hydrogen there, which weakens the binding of $NADP^+$. Nevertheless, the phosphate group can still interact with two residues, Asn41 and Asn44. The presence of these two polar residues enables $NADP^+$ to bind to LXCAR which will use it as cofactor.

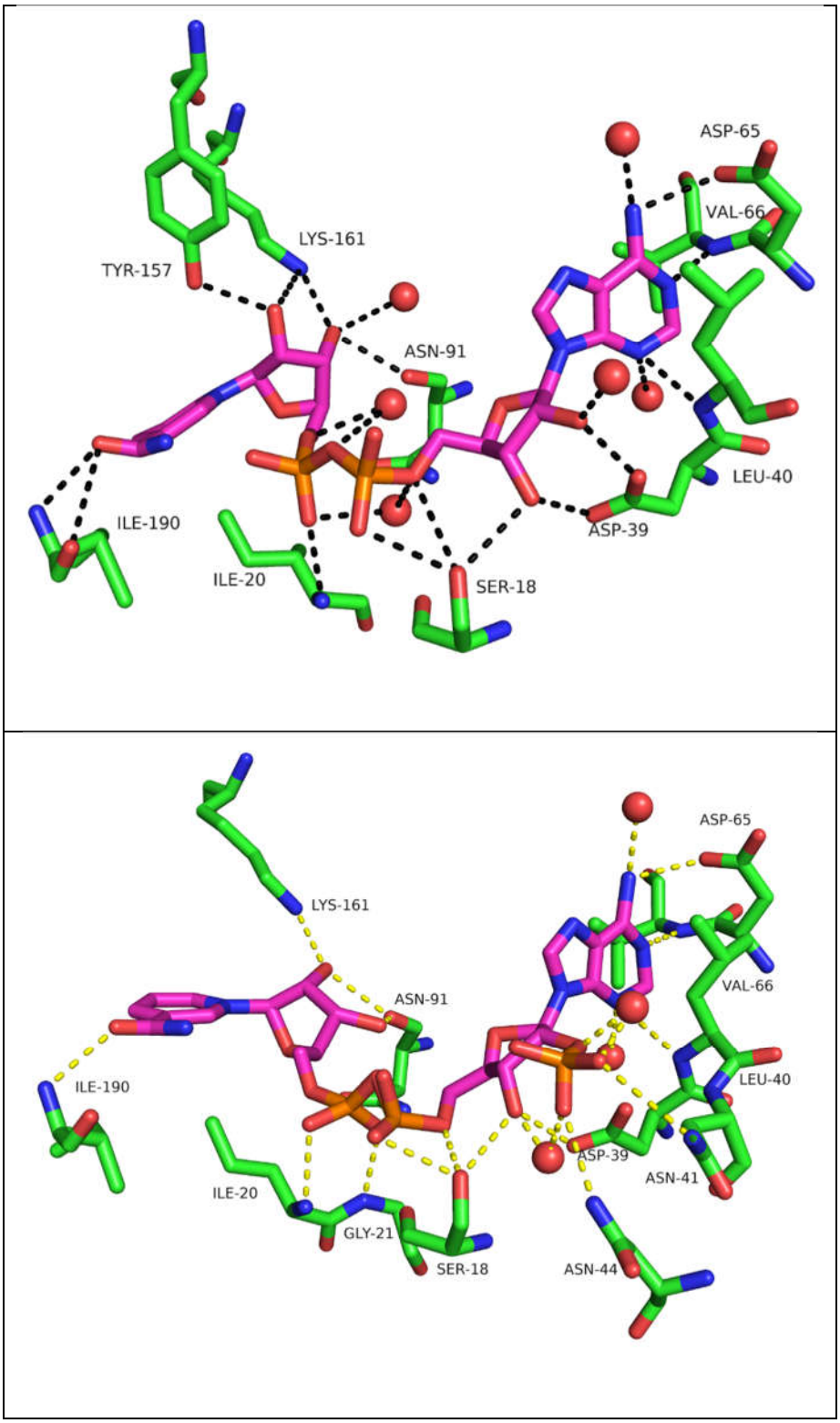


Figure 3.5. Structure of NAD^+ (top) and $NADP^+$ (bottom) docked into LXCAR-S154Y

Catalytic triad

It can be clearly seen that LXCAR conserves the catalytic triad present in numerous SDRs, which includes Ser144, Tyr 157, and Lys 161 (Figure 3.6). Research has demonstrated that the anion form of Tyr hydroxyl group can donate to or accept proton from the substrate, thus playing an essential role in catalysis[33]. The role of Lys161 is to lower the pKa of the Tyr157 hydroxyl, thanks to its electrostatic interaction with the protonated α- amino of Lys161. Nicotinamide rings from NAD^+ were also shown in other SDRs also help reduce Tyr pKa[34]. The deprotonated Tyr157 then acts as a base which deprotonates the hydroxyl group of the substrate. Ser144 is also within hydrogen bonding distance with the substrate OH during the reaction and can therefore stabilize the incipient alkoxide ion by hydrogen bonding as the alcohol is deprotonated. In contrast, during the oxidation reaction, this residue role is polarizing and stabilizing the carbonyl group from substrate. As shown in the above section, beside catalytic capacity, Lys161 participated in the binding of the cofactor, forming an H bond with the 2'-OH of the nicotinamide ribose.

Proton relay system

Various studies have demonstrated that SDRs conserve a proton relay system to shuttle protons from the bulk solvent to a highly conserved Asn residue, usually located on the alpha helix. The proton pathway goes through the side chains of residues involved in the active site including Tyr, Lys, and Asn as well as nicotinamide ribose hydroxyl group and a highly conserved water molecule between Lys and Asn.[14] The X-ray structure of LXCAR clearly shows a proton shuttle system consisting of Tyr157 -OH, the 2'-OH and 3'-OH of NAD^+ ribose, Lys161 -NH3+, and Asn116 -NH2. A water molecule in between Lys161 and Asn116 is also present (Figure 3.7). Here, Asn116 is important in stabilizing the

position of Lys161. When we look at the LXCAR tetramer generated by PISA (in the above section), Asn116 is part of the helix 5 or H5 (Supplemental data, Scheme S3.1), which is right at the dimerization interface of the two LXCAR chains. This area is very water-rich (Supplemental data, Figure S3.4) and can serve as the proton bridge between Lys161 and the bulk solvent. Last, but not least, Asn116 stabilizes the position of the catalytic Lys161. Although Asn116 is not highly conserved in SDRs, crystallographic studies have demonstrated that this residue plays an important role in stabilizing the proton relay system as well as the active site conformation. Sixteen out of twenty SDR structures on PDB possess Asn at homologous positions along with a connecting water the catalytic lysine.[15] While further investigating water molecules adjacent to the connecting one between Asn116 and Lys161, we found a hydrophilic channel that might play a role in shuttling protons in and out of the active site. This includes Ala92, Asn90, Ala119, Val66, Thr38, Gln127, and Asp68. These residues help to transfer protons to the bulk solvent (Supplementary data, Figure S3.5 & S3.6).

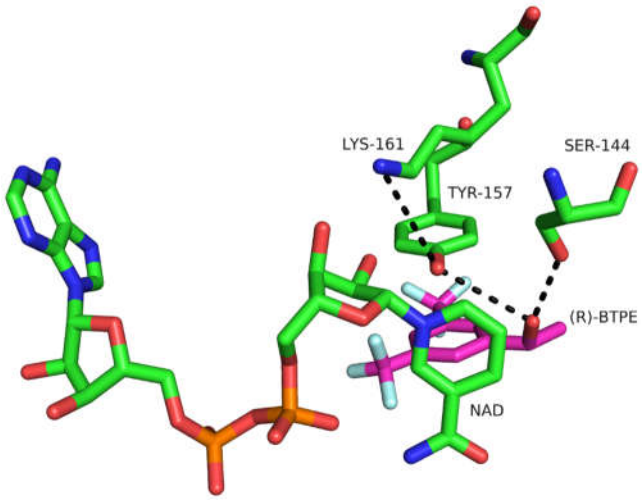


Figure 3.6. Catalytic triad in LXCAR

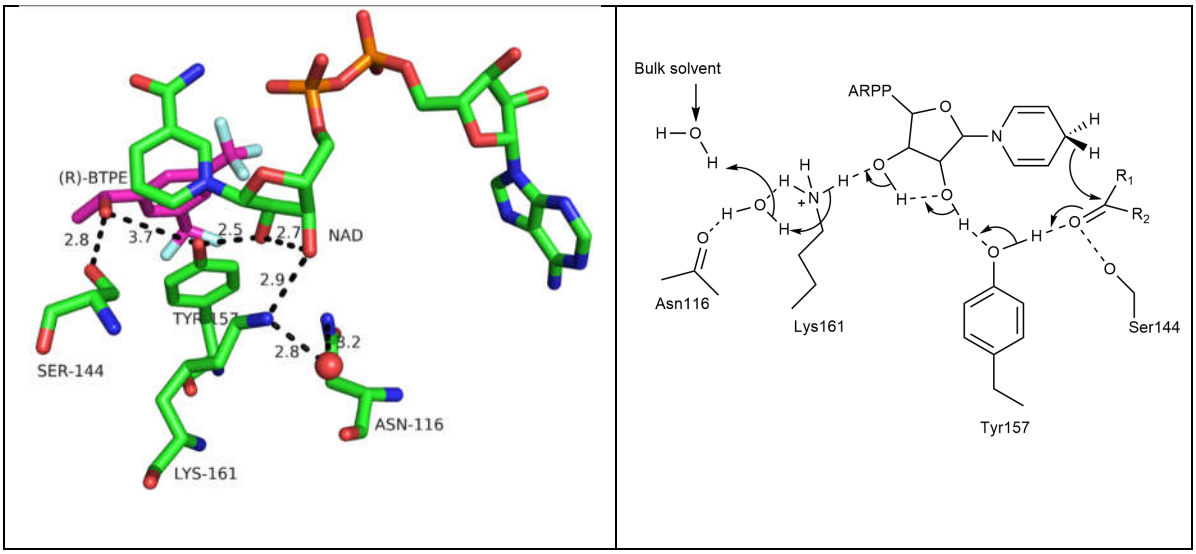


Figure 3.7. Proton relay system in LXCAR; left: shown in stick with H bond length; right: schematic relay mechanism

Conclusion

The crystal structure of the LXCAR-S154Y apoenzyme was obtained at high resolution (1.16 Å), resulting in good crystallographic data. PISA shows that LXCAR is functionally a dimer of dimer in which Mg^{2+} was ligated by 2 chains of the tetramer, stabilizing the quaternary structure. NAD^+ was docked from a homologous model while $NADP^+$ was docked into the structure using Autodock Vina. Like other NAD^+-dependent SDRs, the enzyme utilizes an aspartate residue (Asp39) to bind to 3''-OH of the NAD^+ adenosine ribose. As a phosphate moiety is added to generate a $NADP^+$ molecule, Asp39 loses one hydrogen bond with 2'-OH ribose, which is compensated by the interaction between phosphate group and Asn41 and Asn44. This might be the reason why the protein can react with either NAD^+ or $NADP^+$ as cofactor. (*R*)-BTPE was docked into the apo-enzyme structure of LXCAR-S154Y Autodock vina. The Tyr154 residue was found to play a significant part in the interactions of the trifluoromethyl group. The aromatic ring from Tyr154 allows better interaction to (*R*)-BTPE compared with that of Ser154 from the wild-

type, resulting in better catalysis of LXCAR-S154Y. Last, but not least, a proton relay system was demonstrated in the crystal structure which is very similar to highly conserved systems found in other SDRs. This includes Tyr157, the 2'-OH and 3'OH of NAD^+ ribose, Lys161, and Asn116 and a water molecule in between the last two residues.

Experimental Section

Materials

Kanamycin, isopropyl-β-D-thiogalacto-pyranoside (IPTG), NADPH tetrasodium salt and NADH disodium salt were purchased from Sigma-Aldrich. The DEAE-Sepharose column and Octyl-Sepharose column matrices were purchased from GE Healthcare Life Sciences (USA). All other reagents used in this study were obtained from general commercial sources.

Strains, plasmids, and culture conditions

Recombinant LXCAR-S154Y was synthesized and ligated into the plasmid pET-28a(+) by BIOMATIK (USA). The plasmid (pET-28a(+)-LXCAR-S154Y) was introduced into *E. coli* BL21(DE3) by heat shock before the transformed cells were streaked on LB agar plates supplemented with 50 μg/ml kanamycin and incubated at 37 ºC for 12 h.

Preparation of cell-free extracts of LXCAR-S154Y

A single colony was cultivated overnight at 200 rpm and 37 ºC in 50 mL LB media. Subsequently, 10 mL of this starter culture was added to 1 L of LB media (1 % inoculum, v/v) containing 50 mg kanamycin in a 2 L Erlenmeyer flask, and incubated at 200 rpm and 37 ºC. After the optical density at 600 nm (OD_{600}) reached 0.7 (about 4 h), 0.1 mM IPTG was added, and then cells were shaken at 30 ºC for another 8 h. The cells were harvested by centrifugation at 4000 rpm for 15 mins, resuspended in buffer A (Tris-HCl buffer 50

mM, pH 8.0) and then lysed by ultrasonication (100 W, pulse 1 min) in 4 one-minute intervals (with 10 mins rest in between). The cells debris were removed via centrifugation at 4,000 rpm for 90 mins, to obtain the cell-free extract.

LXCAR-S154Y purification

To obtain high purity LXCAR-S154Y for crystallization, solid ammonium sulfate was added into the crude enzyme extracts to a final saturation of 30%, and the precipitate was collected by centrifugation at 4,000 rpm for 15 min at 4 ºC. Ammonium sulfate was then added to the remaining supernatant to 60% saturation. The precipitate obtained between 30 to 60 % saturation of ammonium sulfate was collected by centrifugation at 4,000 rpm for 15 min and resuspended in buffer A, followed by dialysis against the same buffer at 4 ºC overnight.

Subsequently, the protein was purified using a Low Pressure (LP) chromatography system (Bio-Rad, USA). DEAE-Sepharose (GE Healthcare, USA) was first equilibrated with buffer A and the enzyme loaded at a flow rate of 1.0 mL/min. Buffer A containing 40% 1 M NaCl was used for the elution at 2 mL/min after the non-adsorbed proteins were eluted by buffer A. The collected fractions with carbonyl reductase activity were combined and dialyzed in buffer A containing 0.5 M ammonium sulfate at 4 ºC overnight.

Finally, the dialyzed fraction was loaded on an Octyl-Sepharose column (GE Healthcare, USA) which was pre-equilibrated with buffer A containing 0.5 M ammonium sulfate. The protein was eluted with buffer A. Peak fractions were pooled together and proteins of LXCAR-S154Y were concentrated to approximately 9.1 mg/mL for crystallization.

Measurement of reductase activity and kinetics

All experiments were performed in triplicate. Carbonyl reductase activity was determined with substrate BTAP by monitoring the decrease in absorbance of NADH at 340 nm. The 1.0 mL reaction mixture consisted of 0.3 mM NADH, 10 mM BTAP, 200 mM phosphate buffer (pH 8.0) and 20 μL enzyme solution. One unit of reductase activity was defined as the amount of enzyme that catalyzed the oxidation of 1 μmol NADH per min under the assay conditions. The blank was identical with the assay mixture except without the substrate (BTAP) addition.

Kinetic parameters were determined by using purified recombinant wildtype enyzme (WT-LXCAR) and LXCAR-S154Y with BTAP at concentrations ranging from 1 to 10 mM and 0.3 mM NADH. Data was analyzed by nonlinear regression according to the Michaelis-Menten equation using OriginPro software.

High-throughput screening and Optimization

The initial conditions were found by a high-throughput screening using Qiagen NeXtal Classics I, Classics II, and PEGs I suites. The screen was conducted using a sitting-drop vapor diffusion method with a 10 mg/mL protein solution on three 96-well 2-drop MRC crystallization plates. Formation of crystals was monitored under microscope daily for the first week, every other day for the second week, and weekly since the third week. The crystals shapes and appearances along with their numbers on each well were recorded (Supplemental data, Table S3.1). As a result, the selected condition for optimization was 0.05M-0.2 M $MgCl_2$, 0.1 M HEPES pH 7, and 20-25% (w/v) PEG 3350. Conditions used during optimization were summarized in Supplemental data, Table S3.2.

Protein crystallization and Data collection

Protein crystallization. The protein was crystallized by vapor diffusion in hanging drops by mixing 1 µl of the protein solution (10 mg/ml) with 1 µl of 0.1 M HEPES pH 7, 24% PEG3350, 0.1 M $MgCl_2$, and equilibrated at 293 K over 1 mL of this solution. Crystals appeared after 24-48 h. They were then soaked with solutions containing crystallization solution, 3 mM of either NAD^+ or $NADP^+$ cofactor, and 20% EDG mixture (1:1:1 of ethylene glycol, DMSO, and glycerol) as cryoprotectant before being flash cooled in liquid nitrogen for data collection. The crystals belong to the tetragonal space group P 4_2 2_1 2 with unit cell dimensions of a = 59.29 Å, b = 59.29 Å, c = 118.96 Å, α = β = γ = 90°.

Data collection. Diffraction data were collected at 1 Å, 100 K at the 22-ID beamline of the SER-CAT at the Advanced Photon Source, Argonne National Laboratory. One crystal (0.1 × 0.1 × 0.1 mm) was used to collect data. All data sets were collected with 1 s exposure/0.25°/ frame (rotation of ω). Two data set were collected at 150 mm and 220 mm crystal to detector distances. The data were integrated and reduced using XDS[35] and scaled and merged with the program XSCALE[36].

Structure determination and refinement

Structure determination and refinement. To determine the LXCAR-S154Y structure, a search model was prepared by the MrBump[37] module from CCP4i2, resulting in 5 most homologous possible crystal structures from PDB. Molecular replacement (Phaser-MR-simple one-component interface)[38] in PHENIX was employed using the structure of carbonyl reductase from *Streptomyces coelicolor* (PDB code 5h5x) as the search model. After that, phenix.autobuild[39] was performed to build the structure and initial experimental map. The map was of good quality (initial R_{free} = 0.1680, completeness = 99.9%) and used for further model building in COOT[26] manually. The electron density map calculated at 1

s was well connected. The model was then refined with phenix.refine[40] until R_{free} reached a minimum. Water and ligands were added with COOT.

Structure validation and deposition. The quality of the structure was evaluated with the validation tools included in the programs COOT and MOLPROBITY[41]. Protein quaternary structure analysis was performed using the PISA server (https://www.ebi.ac.uk/pdbe/pisa/). All secondary structure work and figures in this paper was prepared by PyMOL (The PyMOL Molecular Graphics System, Version 1.84, Schrödinger, LLC.). Docking was implemented by Autodock 4.2.1 and Autodock Tools 1.57rc1. Atomic coordinates of LXCAR-S154Y at 1.16 Å (Apo) have been deposited in the PDB and are accessible under PDB ID 6XNB.pdb.

Acknowledgements

This work was supported by the Natural Science Foundation of Zhejiang Province of China under Grant (LGF20B060001); The "13th Five-Year" Chinese Medicine Key Discipline in Zhejiang Province-Chinese Medicine Quality and functional evaluation (2017-XK-A43 to J. Shi) and Provincial first-class (B class) discipline project-Pharmacy. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory, and the University of Georgia X-ray diffraction Core Facility (XRDC). SER-CAT is supported by its member institutions (https://www.ser.aps.anl.gov/www.ser-cat.org/members.html) and equipment grants (S10 RR25528 and S10RR028976) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-

Eng-3. The XRDC is supported by its UGA member groups (http://x-ray.uga.edu/) and an equipment grant (S10 OD021762) from the National Institutes of Health.

References

1. Jin, Y.; Wu, X.; Guan, Y.; Gu, D.; Shen, Y.; Xu, Z.; Wei, X.; Chen, J., Efficacy and safety of aprepitant in the prevention of chemotherapy-induced nausea and vomiting: a pooled analysis. *Supportive Care in Cancer* **2012,** *20* (8), 1815-1822.

2. Wu, C.-E.; Liaw, C.-C., Using aprepitant as secondary antiemetic prophylaxis for cancer patients with cisplatin-induced emesis. *Supportive Care in Cancer* **2012,** *20* (10), 2357-2361.

3. Nakade, S.; Ohno, T.; Kitagawa, J.; Hashimoto, Y.; Katayama, M.; Awata, H.; Kodama, Y.; Miyata, Y., Population pharmacokinetics of aprepitant and dexamethasone in the prevention of chemotherapy-induced nausea and vomiting. *Cancer chemotherapy and pharmacology* **2008,** *63* (1), 75-83.

4. Wan, W.-L.; He, Y.; Guan, M.; Li, X.-L.; Cheng, X.; Wu, Y., Synthesis of the major isomers of Aprepitant and Fosaprepitant. *Chinese Chemical Letters* **2013,** *24* (12), 1118-1122.

5. Zhao, M. M.; McNamara, J. M.; Ho, G.-J.; Emerson, K. M.; Song, Z. J.; Tschaen, D. M.; Brands, K. M.; Dolling, U.-H.; Grabowski, E. J.; Reider, P. J., Practical asymmetric synthesis of aprepitant, a potent human NK-1 receptor antagonist, via a stereoselective Lewis acid-catalyzed trans acetalization reaction. *The Journal of organic chemistry* **2002,** *67* (19), 6743-6747.

6. Brands, K. M.; Payack, J. F.; Rosen, J. D.; Nelson, T. D.; Candelario, A.; Huffman, M. A.; Zhao, M. M.; Li, J.; Craig, B.; Song, Z. J., Efficient synthesis of NK1 receptor antagonist aprepitant using a crystallization-induced diastereoselective transformation. *Journal of the American Chemical Society* **2003,** *125* (8), 2129-2135.

7. Wang, P.; Cai, J.-B.; Ouyang, Q.; He, J.-Y.; Su, H.-Z., Asymmetric biocatalytic reduction of 3,5-bis(trifluoromethyl)acetophenone to (1*R*)-[3,5-bis(trifluoromethyl)phenyl]ethanol using whole cells of newly isolated *Leifsonia xyli* HS0904. *Applied microbiology and biotechnology* **2011,** *90* (6), 1897-1904.

8. Gräff, M.; Buchholz, P. C.; Stockinger, P.; Bommarius, B.; Bommarius, A. S.; Pleiss, J., The Short-chain Dehydrogenase/Reductase Engineering Database (SDRED): A classification and analysis system for a highly diverse enzyme family. *Proteins: Structure, Function, and Bioinformatics* **2019,** *87* (6), 443-451.

9. Stavrinides, A. K.; Tatsis, E. C.; Dang, T. T.; Caputi, L.; Stevenson, C. E.; Lawson, D. M.; Schneider, B.; O'Connor, S. E., Discovery of a Short-Chain Dehydrogenase from *Catharanthus roseus* that Produces a New Monoterpene Indole Alkaloid. *ChemBioChem* **2018,** *19* (9), 940-948.

10. Prelog, V., Specification of the stereospecificity of some oxido-reductases by diamond lattice sections. *Pure Appl. Chem* **1964,** *9* (1), 119-130.

11. Lygidakis, A.; Karuppiah, V.; Hoeven, R.; Ní Cheallaigh, A.; Leys, D.; Gardiner, J. M.; Toogood, H. S.; Scrutton, N. S., Pinpointing a Mechanistic Switch Between Ketoreduction and “Ene” Reduction in Short-Chain Dehydrogenases/Reductases. *Angewandte Chemie* **2016,** *128* (33), 9748-9752.

12. Kallberg, Y.; Oppermann, U.; Jörnvall, H.; Persson, B., Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes. *Protein Science* **2002,** *11* (3), 636-641.

13. Lerchner, A.; Jarasch, A.; Meining, W.; Schiefner, A.; Skerra, A., Crystallographic analysis and structure-guided engineering of NADPH-dependent *Ralstonia* sp. Alcohol dehydrogenase toward NADH cosubstrate specificity. *Biotechnology and bioengineering* **2013,** *110* (11), 2803-2814.

14. Filling, C.; Berndt, K. D.; Benach, J.; Knapp, S.; Prozorovski, T.; Nordling, E.; Ladenstein, R.; Jörnvall, H.; Oppermann, U., Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry* **2002,** *277* (28), 25677-25684.

15. Kubota, K.; Nagata, K.; Okai, M.; Miyazono, K.-i.; Soemphol, W.; Ohtsuka, J.; Yamamura, A.; Saichana, N.; Toyama, H.; Matsushita, K., The crystal structure of L-sorbose reductase from *gluconobacter frateurii* complexed with NADPH and L-sorbose. *Journal of molecular biology* **2011,** *407* (4), 543-555.

16. Kristan, K.; Stojan, J.; Adamski, J.; Rižner, T. L., Rational design of novel mutants of fungal 17β-hydroxysteroid dehydrogenase. *Journal of biotechnology* **2007,** *129* (1), 123-130.

17. Kallberg, Y.; Oppermann, U.; Jörnvall, H.; Persson, B., Short-chain dehydrogenases/reductases (SDRs) Coenzyme-based functional assignments in completed genomes. *European Journal of Biochemistry* **2002,** *269* (18), 4409-4417.

18. Wang, N.-Q.; Sun, J.; Huang, J.; Wang, P., Cloning, expression, and directed evolution of carbonyl reductase from *Leifsonia xyli* HS0904 with enhanced catalytic efficiency. *Applied microbiology and biotechnology* **2014,** *98* (20), 8591-8601.

19. Wang, N.; Huang, J.; Luo, H.; Wang, P.; Li, J., Purification and characterization of a new carbonyl reductase from *Leifsonia xyli* HS0904 involved in stereoselective reduction of 3,5-bis(trifluoromethyl)acetophenone. *Journal of Molecular Catalysis B: Enzymatic* **2013,** *92*, 1-6.

20. Büsing, I.; Höffken, H. W.; Breuer, M.; Wöhlbrand, L.; Hauer, B.; Rabus, R., Molecular Genetic and Crystal Structural Analysis of 1-(4-Hydroxyphenyl)-Ethanol Dehydrogenase from '*Aromatoleum aromaticum*' EbN1. *Journal of Molecular Microbiology and Biotechnology* **2015,** *25* (5), 327-339.

21. Li, T.-B.; Zhao, F.-J.; Liu, Z.; Jin, Y.; Liu, Y.; Pei, X.-Q.; Zhang, Z.-G.; Wang, G.; Wu, Z.-L., Structure-guided engineering of ChKRED20 from *Chryseobacterium* sp. CA49 for asymmetric reduction of aryl ketoesters. *Enzyme and microbial technology* **2019,** *125*, 29-36.

22. Paithankar, K. S.; Feller, C.; Kuettner, E. B.; Keim, A.; Grunow, M.; Sträter, N., Cosubstrate-induced dynamics of D-3-hydroxybutyrate dehydrogenase from *Pseudomonas putida*. *The FEBS journal* **2007,** *274* (21), 5767-5779.

23. Niefind, K.; Müller, J.; Riebel, B.; Hummel, W.; Schomburg, D., The crystal structure of R-specific alcohol dehydrogenase from *Lactobacillus brevis* suggests the structural basis of its metal dependency. *Journal of molecular biology* **2003,** *327* (2), 317-328.

24. Schlieben, N. H.; Niefind, K.; Müller, J.; Riebel, B.; Hummel, W.; Schomburg, D., Atomic resolution structures of R-specific alcohol dehydrogenase from *Lactobacillus*

brevis provide the structural bases of its substrate and cosubstrate specificity. *Journal of molecular biology* **2005,** *349* (4), 801-813.

25. Graninger, M.; Nidetzky, B.; Heinrichs, D. E.; Whitfield, C.; Messner, P., Characterization of dTDP-4-dehydrorhamnose 3, 5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *Journal of Biological Chemistry* **1999,** *274* (35), 25069-25077.

26. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (4), 486-501.

27. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **2009,** *30* (16), 2785-2791.

28. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010,** *31* (2), 455-461.

29. Pennacchio, A.; Giordano, A.; Esposito, L.; Langella, E.; Rossi, M.; Raia, C. A., Insight into the stereospecificity of short-chain *Thermus thermophilus* alcohol dehydrogenase showing pro-*S* hydride transfer and prelog enantioselectivity. *Protein and peptide letters* **2010,** *17* (4), 437-443.

30. Esterhuysen, C.; Heßelmann, A.; Clark, T., Trifluoromethyl: an amphiphilic noncovalent bonding partner. *ChemPhysChem* **2017,** *18* (7), 772-784.

31. Hörer, S.; Stoop, J.; Mooibroek, H.; Baumann, U.; Sassoon, J., The crystallographic structure of the mannitol 2-dehydrogenase $NADP^+$ binary complex from Agaricus bisporus. *Journal of Biological Chemistry* **2001,** *276* (29), 27555-27561.

32. Otagiri, M.; Kurisu, G.; Ui, S.; Takusagawa, Y.; Ohkuma, M.; Kudo, T.; Kusunoki, M., Crystal structure of meso-2,3-butanediol dehydrogenase in a complex with NAD^+ and inhibitor mercaptoethanol at 1.7 Å resolution for understanding of chiral substrate recognition mechanisms. *The Journal of Biochemistry* **2001,** *129* (2), 205-208.

33. Kavanagh, K.; Jörnvall, H.; Persson, B.; Oppermann, U., Medium-and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences* **2008,** *65* (24), 3895-3906.

34. Koumanov, A.; Benach, J.; Atrian, S.; Gonzàlez-Duarte, R.; Karshikoff, A.; Ladenstein, R., The catalytic mechanism of *Drosophila* alcohol dehydrogenase: evidence for a proton relay modulated by the coupled ionization of the active site lysine/tyrosine pair and a NAD^+ ribose OH switch. *Proteins: Structure, Function, and Bioinformatics* **2003,** *51* (2), 289-298.

35. Kabsch, W., Xds. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 125-132.

36. Kabsch, W., Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 133-144.

37. Keegan, R. M.; Winn, M. D., MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **2008,** *64* (1), 119-124.

38. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (4), 658-674.

39. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L.-W.; Read, R. J.; Adams, P. D., Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography* **2008,** *64* (1), 61-69.

40. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* **2012,** *68* (4), 352-367.

41. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **2018,** *27* (1), 293-315.

CHAPTER 4 - CONCLUSIONS

Alcohol dehydrogenase (ADH) family has been studied heavily in recent years owning to their potential application of pharmaceutical syntheses of extremely important intermediates. The oxidation direction requires a net removal of two hydrogen atoms from the alcohol. Thus, this reaction occurs by coupling transfer of a hydride (H^-) to cofactor and abstraction of a proton (H^+) which in turn will be shuttled by a proton relay system to the solvent bulk. More specifically, the substrate oxidation includes four steps, namely i) alcohol binding, ii) proton abstraction, iii) hydride transfer, and iv) ketone and cofactor release. Among many ADHs, *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase (TeSADH) is among the most promising enzyme thanks to its ability to work in harsh conditions like high temperature (up to 70 ºC) and in the presence of several organic solvents. Furthermore, it is very stable after being immobilized on solid support such as Eupergit-C resin, remaining consistently active for 30 days at 37 ºC. Like many of the medium chain dehydrogenases/reductases (MDRs), TeSADH is a metalloenzyme and more specifically a zinc-dependent homotetrameric protein. Research of zinc-dependent enzymes has reemerged recently due to the central role of this metal ion in numerous catalytic mechanisms. However, there are very little information about the zinc intermediates states changing during the catalysis of zinc-dependent enzymes due to the silence of Zn^{2+} for many spectroscopic techniques, including EPR and optical

spectroscopy. The dynamics of zinc intermediate have remained a subject of debate between biochemists. While method like X-ray absorption spectroscopy (XAS) can be used to investigate the coordination sphere of Zn^{2+}, this method results are not conclusive due to its limitation. including heterogeneity of zinc ligands in the structures of the enzyme−NADP. X-ray crystallography offers an alternative method that gives a more intuitive picture of the enzyme active site during catalysis. One of our crystal structure revealed a pentacoordinate Zn^{2+} which is involved in Cys-27, His-59, Glu-60, Asp-150 and a water molecule at location previously occupied by substrates. The structure showed a pre-pentacoordinate zinc state with Cys-27, His-59, Asp-150, and several substrates in the first coordination sphere while the Glu-60 is in the distant second coordination sphere. Unfortunate, we did not obtain a pentacoordinate zinc complex with substrate ligand. In the future, new techniques are required to prove the pentacoordinate transition state with alcohol binding. These results support one proposed model as follow. First, a water binds to the metal ion to form a pentacoordinate zinc intermediate before being displaced by an alcohol. Then, zinc coordination is in tetrahedral geometry since the alcohol is still on the second coordination sphere. Subsequently, the substrate covalently bonds to Zn^{2+}, giving rise to a new pentacoordinate complex before Glu-60 moves away, resulting in a tetracoordinated complex. This is followed by ketone and cofactor releases. Our results invalidate the model in which alcohol and water molecules simultaneously coordinate to zinc before Glu-60 is displaced. By comparison, the catalytic mechanism of SADH from *T. ethanolicus* bacteria is different from that of mammalian counterparts, for instance horse liver alcohol dehydrogenase (HLADH). Whereas pentacoordinate zinc has been only demonstrated in other enzyme family, our crystal structures were the first ever with

observed pentacoordinate complex in ADH family, as well as the first ternary structure including both substrates and cofactors for TeSADH. Last but not least, our structures reveal a proton shuttle system from Ser-39 via the 2'-and 3'-OH of NADP ribose to His-42 which were not demonstrated in previous ones.

Although X-ray crystallography is a very powerful method, it has its own disadvantages. Firstly, finding a good condition and crystallizing a protein to obtain well-diffracted crystals are very challenging. Secondly, some enzymes either takes a long time to be crystallized or their crystallizations are not always reproducible. Both issues happen to TeSADH. More importantly, X-ray diffraction must be performed at very low temperature (100 K) to minimize radiation damage. The three-dimensional structures of protein at this temperature might be significantly different from those of at physiological temperature. Last but not least, as the enzyme molecules must arrange in their crystal lattice, they might occasionally block the channel in which the substrate molecules travel to the active site. This in fact happened to the second protein in my research project, *Leifsonia xyli* carbonyl reductase (LXCAR) S154Y mutant. Even though we have tried numerous modifications like increasing soaking concentration and incubation time, the obtain X-ray structure of apo enzyme with no bound substrates or cofactors. To circumvent this problem, we employed docking program including Autodock Tools and Autodock vina to incorporate ligands into the first ever LXCAR- S154Y crystal structure. Nevertheless, our results reveal interesting aspects of this mutant. Firstly, LXCAR dimerizes thanks to the presence of Mg^{2+} which is coordinated by two C-terminal Gln-251 from two chains and four coplanar water oxygen atoms in hexacoordinate octahedral complex. This helps to hold the two subunits together and stabilize the resulting dimer. Docking separately of

NAD^+ and $NADP^+$ explain why LXCAR can use either of them as cofactor. Like other NAD^+-dependent SDRs, this enzyme uses an aspartate residue (Asp39) to form hydrogen bond with to 3''-OH of the NAD^+ adenosine ribose. As a phosphate moiety is added to generate a $NADP^+$ molecule, Asp39 loses one hydrogen bond with 2'-OH ribose, which is compensated by the interactions between phosphate group and Asn41 and Asn44. Docking of (*R*)-BTPE showed that Tyr154 residue plays a significant role in the interaction of the trifluoromethyl group. The aromatic ring from Tyr154 allows better interaction to (*R*)-BTPE compared with that of Ser154 from the wild-type, resulting in better catalysis of LXCAR-S154Y. Finally, a proton relay system was demonstrated in the crystal structure which is very similar to highly conserved systems found in other SDRs. This includes Tyr157, the 2'-OH and 3'OH of NAD^+ ribose, Lys161, and Asn116 and a water molecule in between the last two residues. Because of the extreme significance of this enzyme and its mutant in the production of (*R*)-BTPE, essential intermediates for syntheses of two drugs Aprepitant and Fosaprepitant, this is a need of using alternative method to obtain three-dimensional structure of LXCAR with bound ligands. Co-crystallization and Cryogenic electron microscopy (Cryo-EM) are the methods of recommendation. While the former may solve the problem, it requires independent high-throughput screening for each new set of ligands, which is very time-consuming and not cost-effective. Therefore, in my opinion, Cryo-EM should be preferable for several reasons. Firstly, it does not require crystallization of proteins. Secondly, as enzyme molecules do not pack in a crystal lattice, this method circumvent the issue of ligand being blocked from approaching the active site. With great advancement of Cryo-EM in recently, obtaining high resolution of LXCAR ternary structures containing various substrates and cofactors is feasible.

APPENDIX A

CHAPTER 2 SUPPORTING INFORMATION

TABLE OF CONTENTS

Page

Table S2.1. Optimization of KCl and PEG 3350 concentrations using HEPES buffer pH 7.5

	0.45 M KCl	0.6 M KCl	0.75 M KCl
8% PEG 3350	Clear	Clear	Clear
10% PEG 3350	Clear	Clear	Clear
12% PEG 3350	Crystals	Crystals	Small crystals
15% PEG 3350	Light precipitate	Light precipitate	Light precipitate

Optimization for TeSADH-I86A crystallization: Formation of crystals was monitored under microscope daily for the first week, every other day for the second week, and weekly since the third week. The crystals shapes and appearances along with their numbers on each well were recorded. It was found that at PEG 3350 concentration of 12% and KCl concentration of 0.6 M (Supplemental date, Table S1). It is clear from the table that nucleation of the protein did not occur at PEG 3350 concentrations lower than 12%. At concentration for PEG larger than 12% (15% or higher), light precipitates were seen, which indicated supersaturation. In conclusion, 12% is optimal concentration for PEG 3350. The change in KCl concentration, however, did not result in such obvious effect. The crystals formed in 0.6 M KCl solution only slightly larger than other crystals. Due to this, the KCl concentration was varied while performing the pH screen to see if a more definite trend could be observed. Using the 12% PEG 3250 as a constant, the pH screening was conducted with 50 mM HEPES pH ranging from 7.2-7.5. The results showed that crystals of a similar distorted octahedron structure formed at pH values of 7.2-7.5 and a KCl concentration of

0.6 M. The dimensions of the crystal are roughly 100 x 100 x 100 µm with the crystals formed at pH 7.5 being the largest. Moreover, their shape is more octahedron-like. Crystals at 7.2-7.4, which were distorted octahedron and smaller in size, was formed in presence of light precipitate.

Figure S2.1. Crystal of TeSADH-C295A

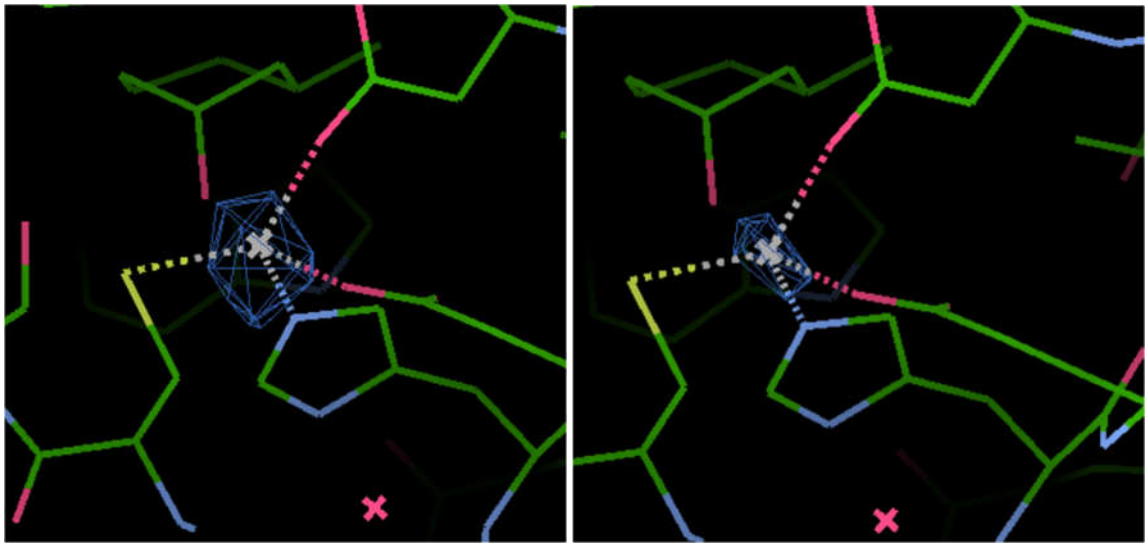


Figure S2.2. Unusual map coefficient (cmapcoeff) by in ccp4i2 package24 overlaid with the model from MR-SAD in chain B at contour σ = 7 (left) and σ = 6 (right)

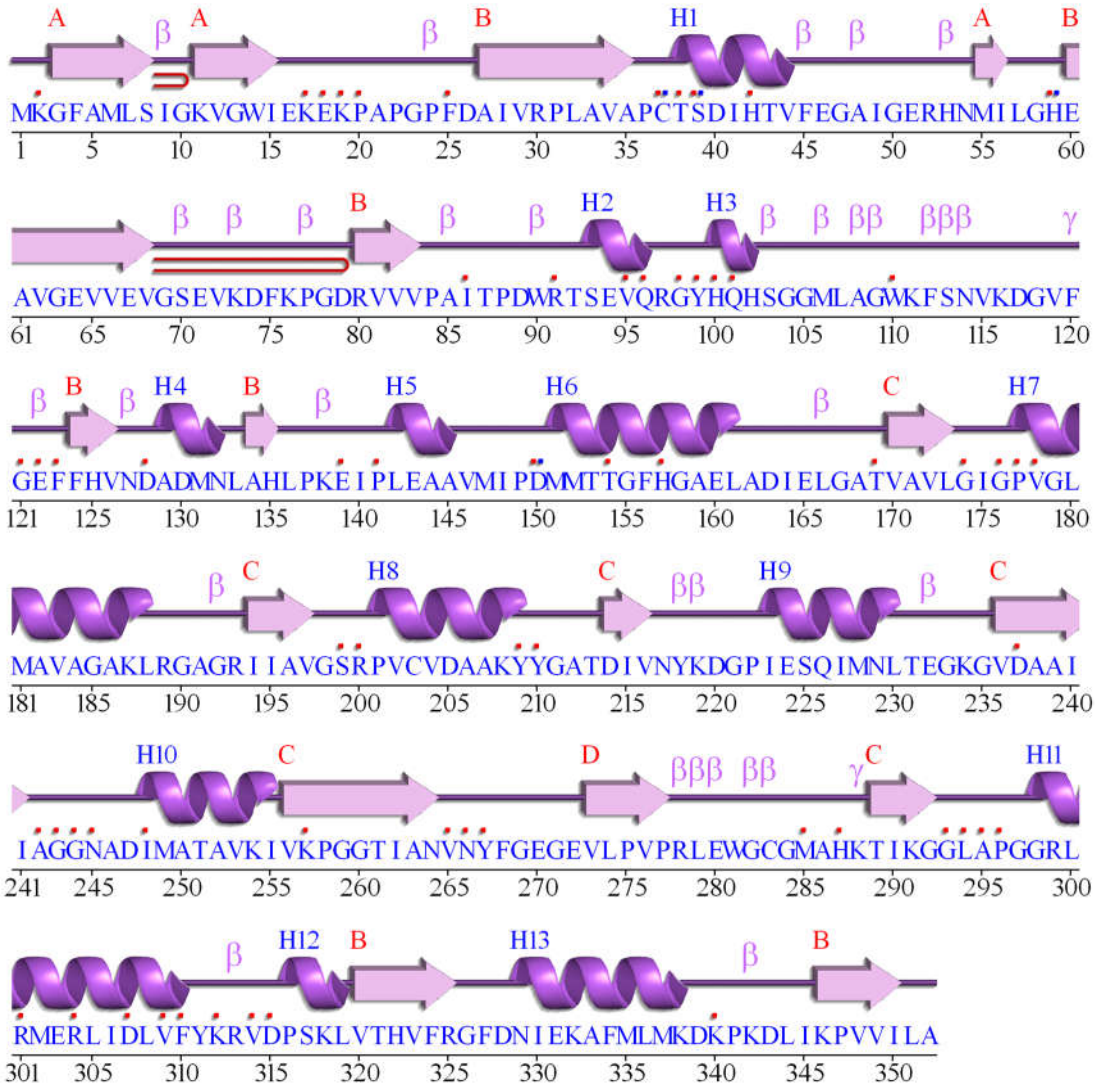


Figure S2.3. Secondary structure of TeSADH-C295A

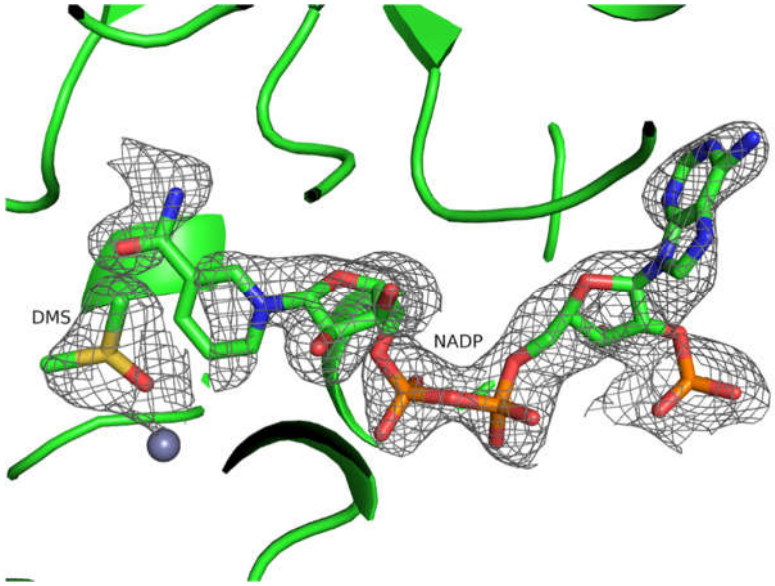


Figure S2.4. Simulated-annealing OMIT map for NADP and DMSO molecules in 7JNQ at chain D (σ = 3)

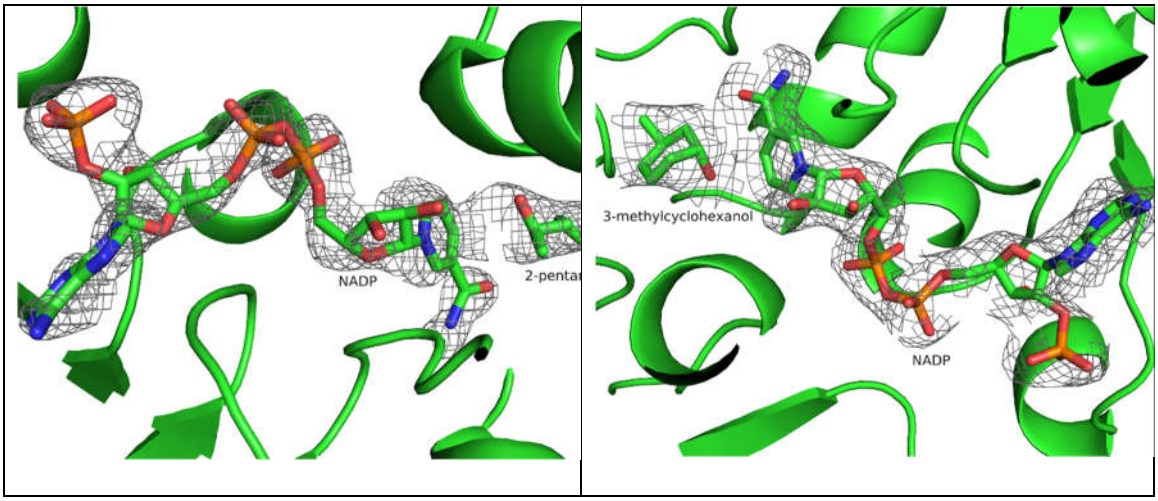


Figure S2.5. Simulated-annealing OMIT maps for NADP and 2-pentanol in 7JNU (left) or 3-methylcylohexanol (right) in 7JNS (σ = 3)

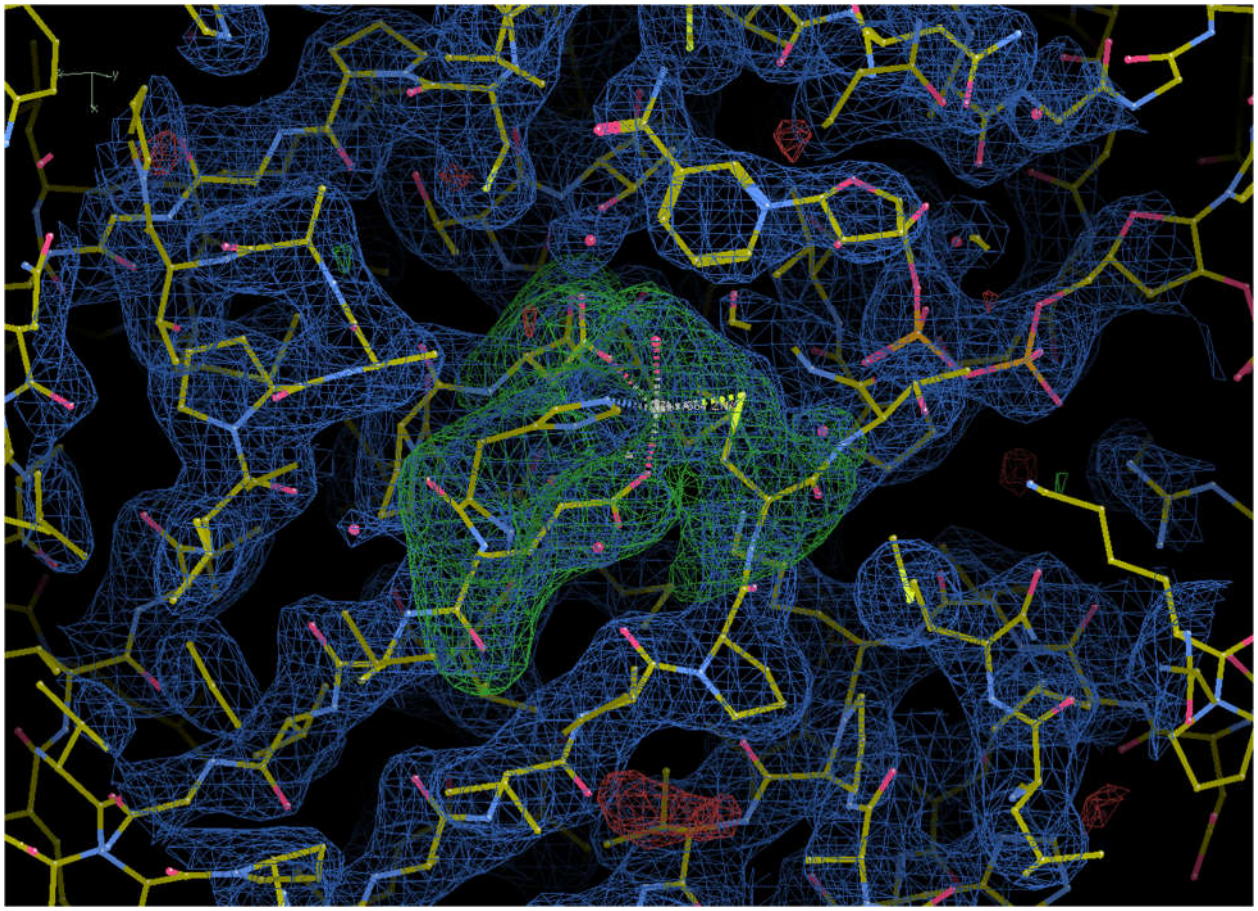


Figure S2.6. Sim omit map of the pentacoordinate zinc in I86A-3-methylcyclohexanol structure (presented by coot)

APPENDIX B

CHAPTER 3 SUPPORTING INFORMATION

TABLE OF CONTENTS

Page

Table S3.1. Summary of conditions which resulted in best crystals from high-throughput screening experiment of LXCAR-S154Y

Salt	pH	Buffer	Crystallizing agent (w/v)	Crystal appearance	Number of crystals *
0.2 M $Ca(AcO^-)_2$	-	-	20% **PEG 3350**	**bipyramidal**	2
0.1 M DL-Malic acid	7	-	12% **PEG 3350**	rod	2
1.1 M Sodium malonate	7	**0.1 M HEPES**	0.5% (v/v) jeffamine ED-2001	cubic	3
1 M Succinic acid	7	**0.1 M HEPES**	1% PEG 2000 MME	triangular prism	2
0.02 $(NH_4)_2SO_4$	7	**0.1 M HEPES**	0.5% PEG 8000	bipyramidal	2
0.2 M $MgCl_2$	7.5	**0.1 M HEPES**	22% sodium polyacrylate 5100	**bipyramidal**	2
0.2 M $MgCl_2$	5.5	0.1 M Bis-Tris	**25% PEG 3350**	**bipyramidal**	2
0.2 M $MgCl_2$	6.5	0.1 M Bis-Tris	**25% PEG 3350**	**bipyramidal**	2
0.2 M $MgCl_2$	8.5	0.1 M Tris	**25% PEG 3350**	**bipyramidal**	1
0.2 M $MgCl_2$	7.5	**0.1 M HEPES**	**25% PEG 3350**	**bipyramidal**	1

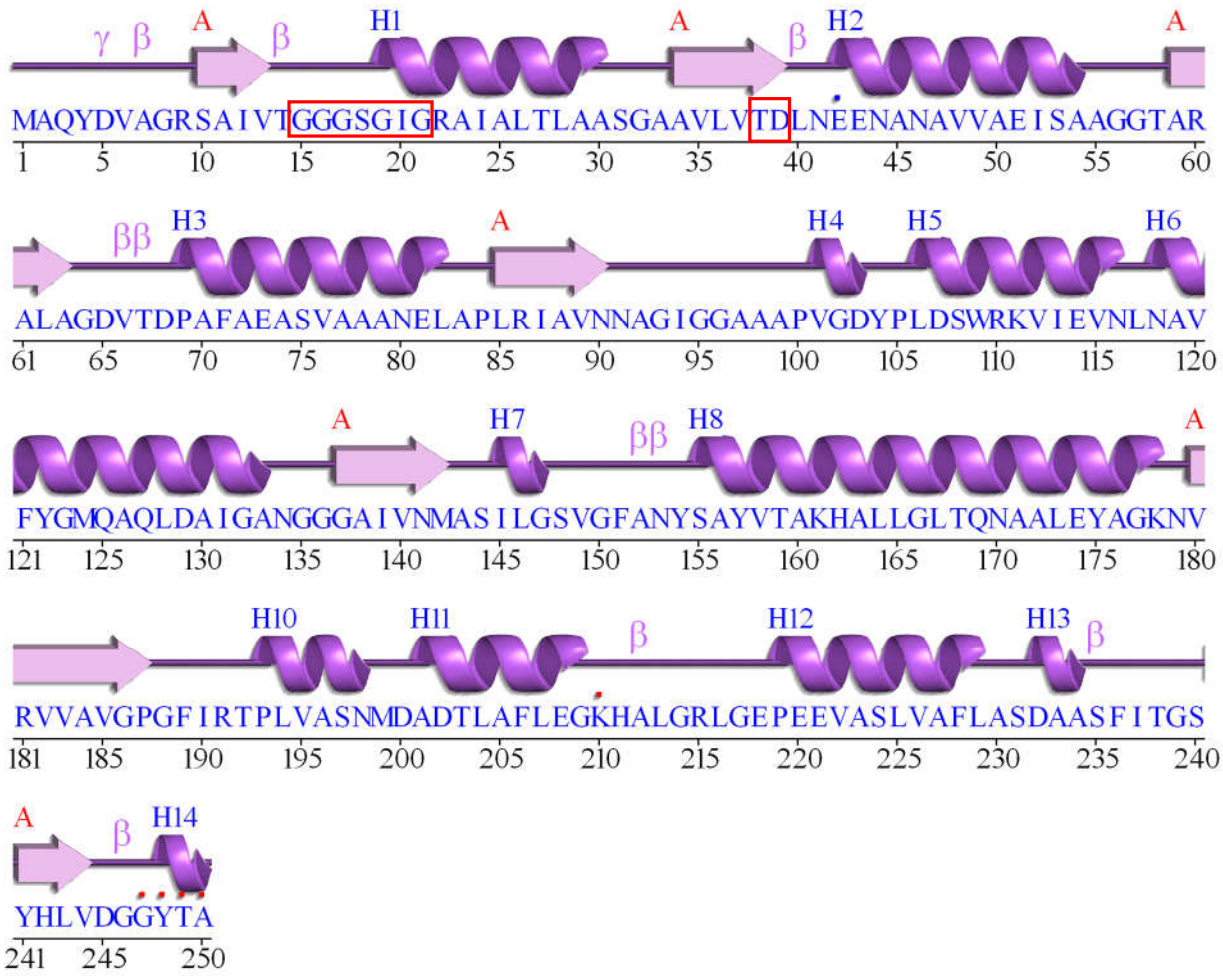
* Number crystal after a month, which had unchanged since then, were monitored weekly

Table S3.2. Condition for crystallization was optimized using a hanging drop vapor diffusion method

PEG (w/v) / $MgCl_2$	20%	21%	22%	23%	24%	25%
0.05 M	-	-	-	-	-	-
0.1 M	-	-	-	3	10 crystals	-
0.2 M	-	-	-	-	-	-

Condition: 2 µL drop (mixing 1:1 protein and well solutions) with 1,000 µL well solution on a Hampton 24-well crystallization plate. The finally selected condition was PEG3350, 0.1 M $MgCl_2$.

"-": means no crystals



Scheme S3.1. Structural summary of LXCAR-S154Y created by pdbsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html), the regions to determine cofactor dependency are put in red boxes

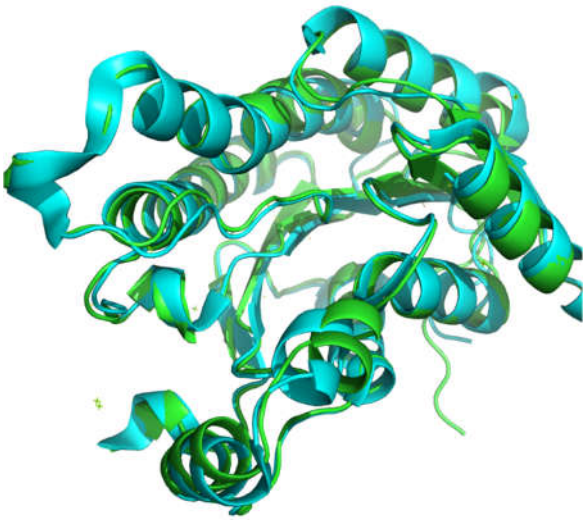


Figure S3.1. LXCAR overlay with 5H5X monomer

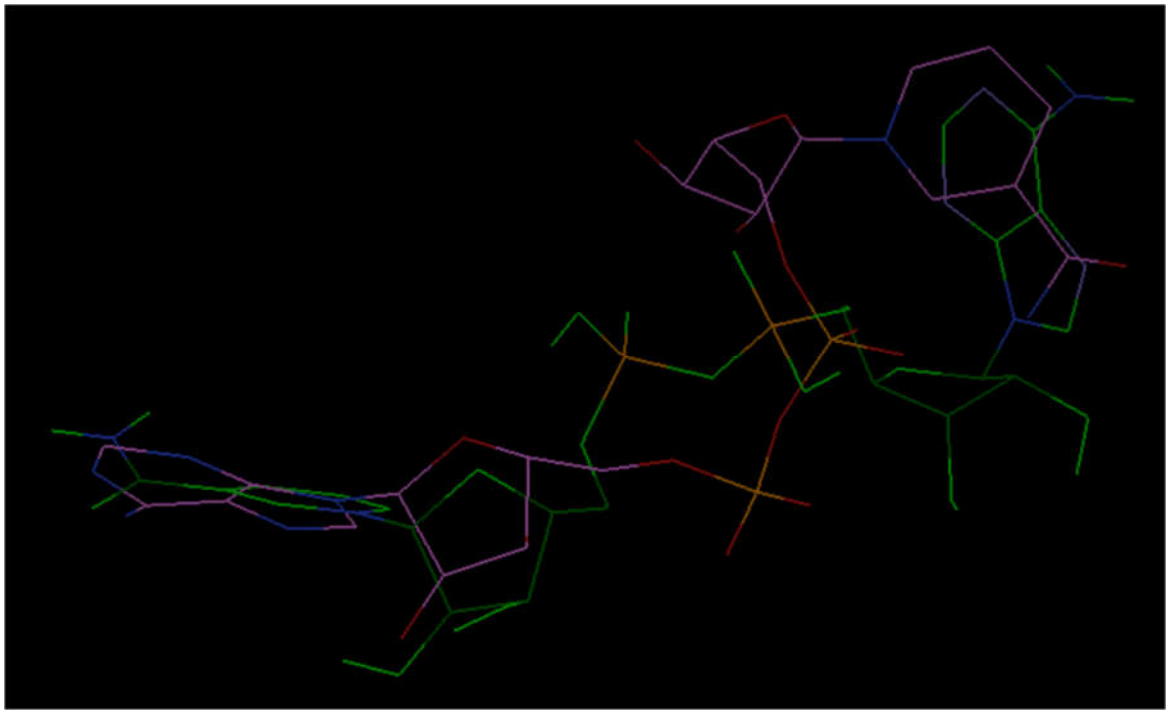
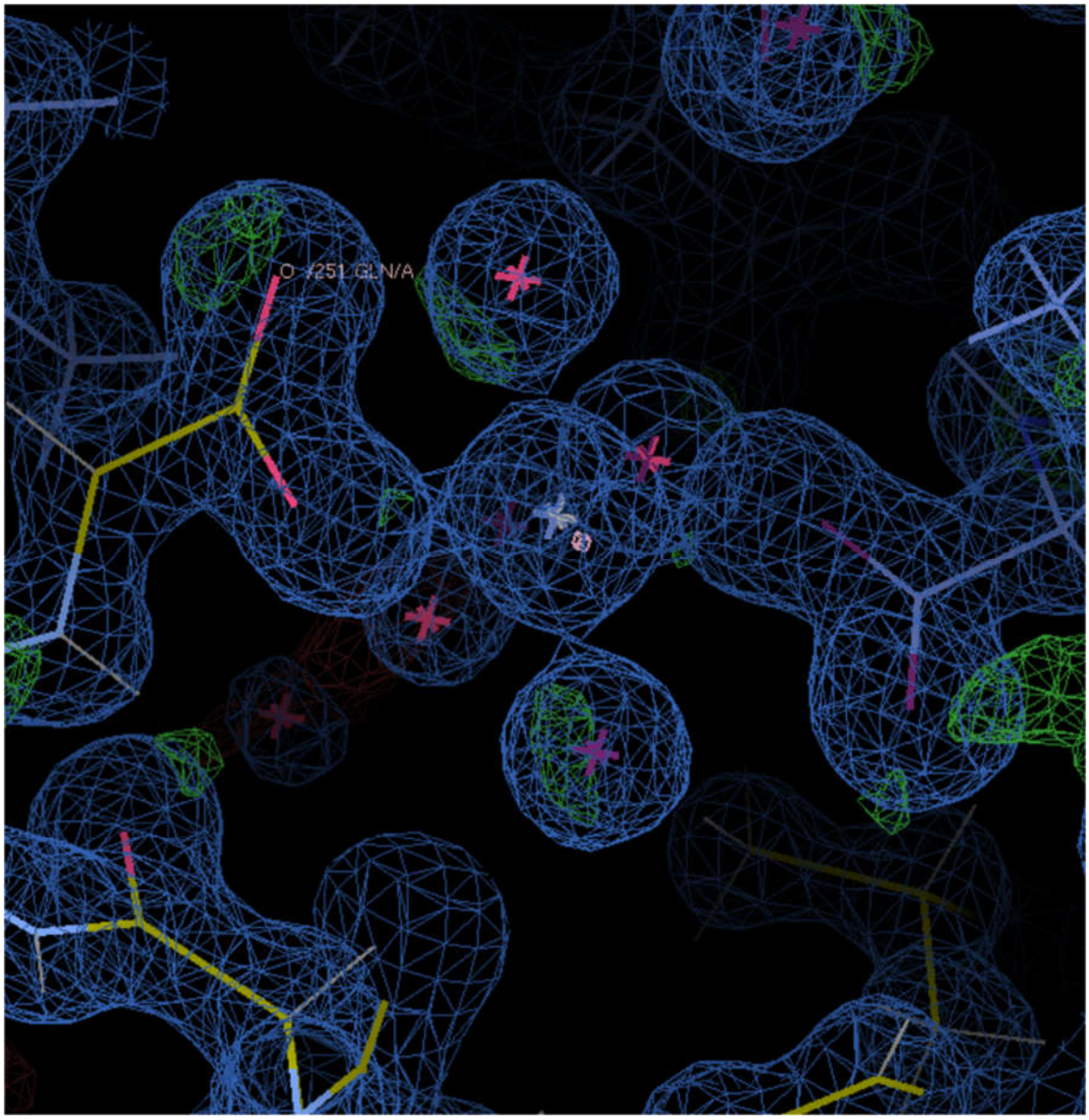


Figure S3.2. Positioning of NAD^+ from 5h5x (in green) versus redocked NAD^+ by Autodock tools (in purple)


O 251 GLN/A

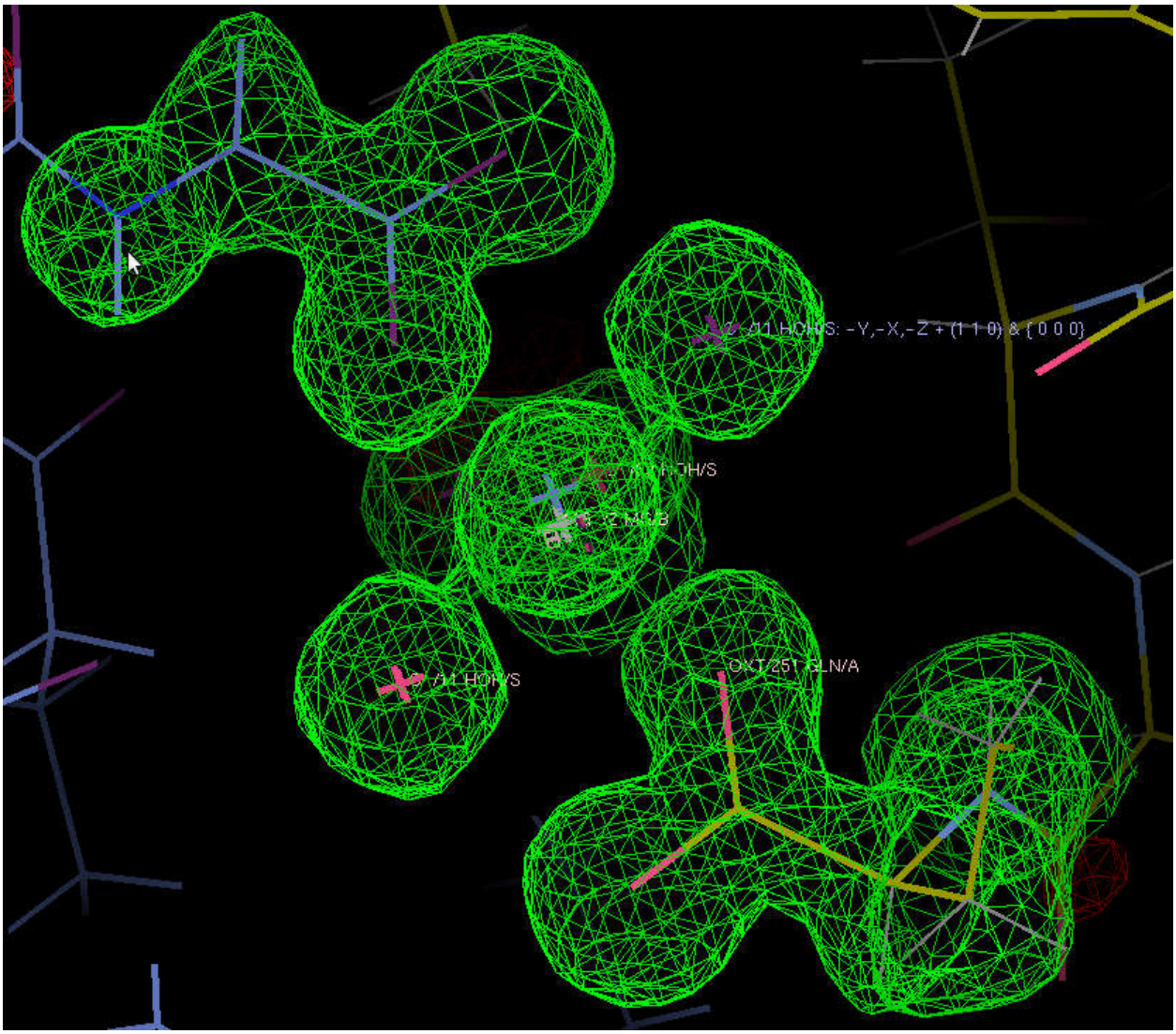


Figure S3.3. Electron density $2F_o$-F_c (top) and F_o-F_c (bottom) of Mg^{2+} along with four coordinated water molecules and Gln-251 in LXCAR

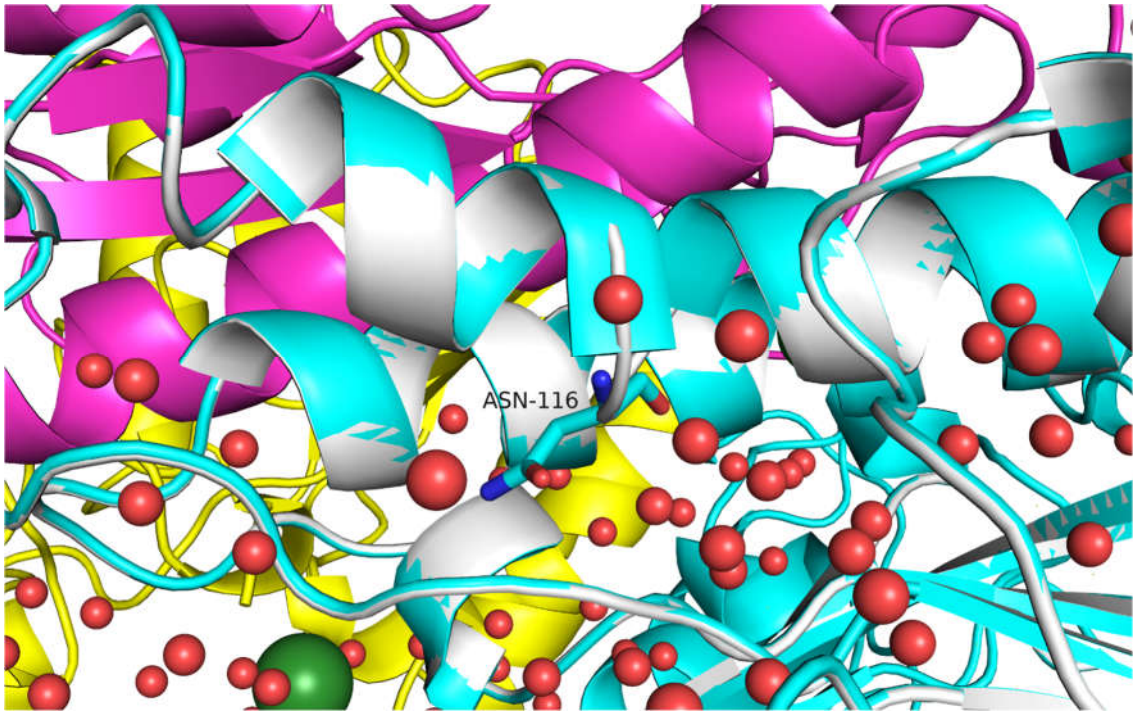


Figure S3.4. Water-rich channel near helix 5 (H5) of LXCAR which is part of the interface between 2 chains in LXCAR tetramer (cyan and violet)

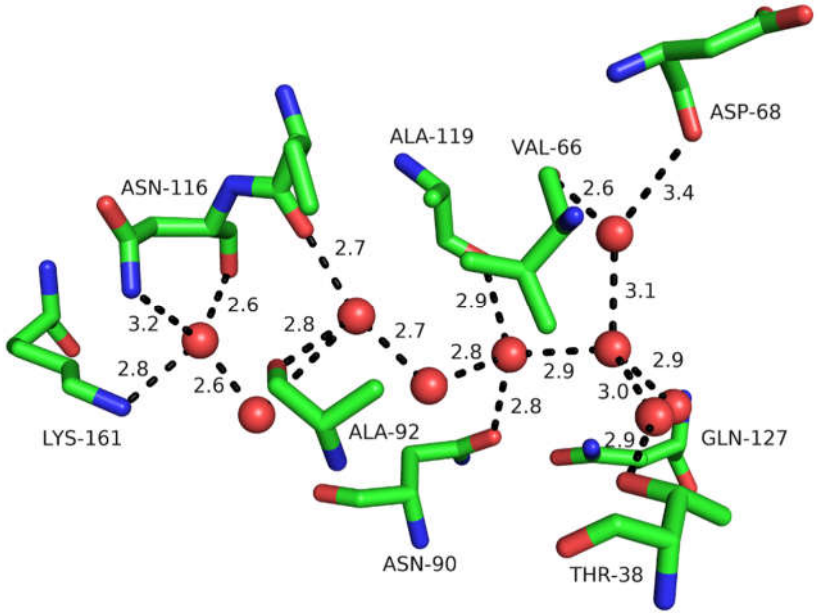


Figure S3.5. Hydrophilic and water-rich channel that shuttles proton from Asn-116 to the

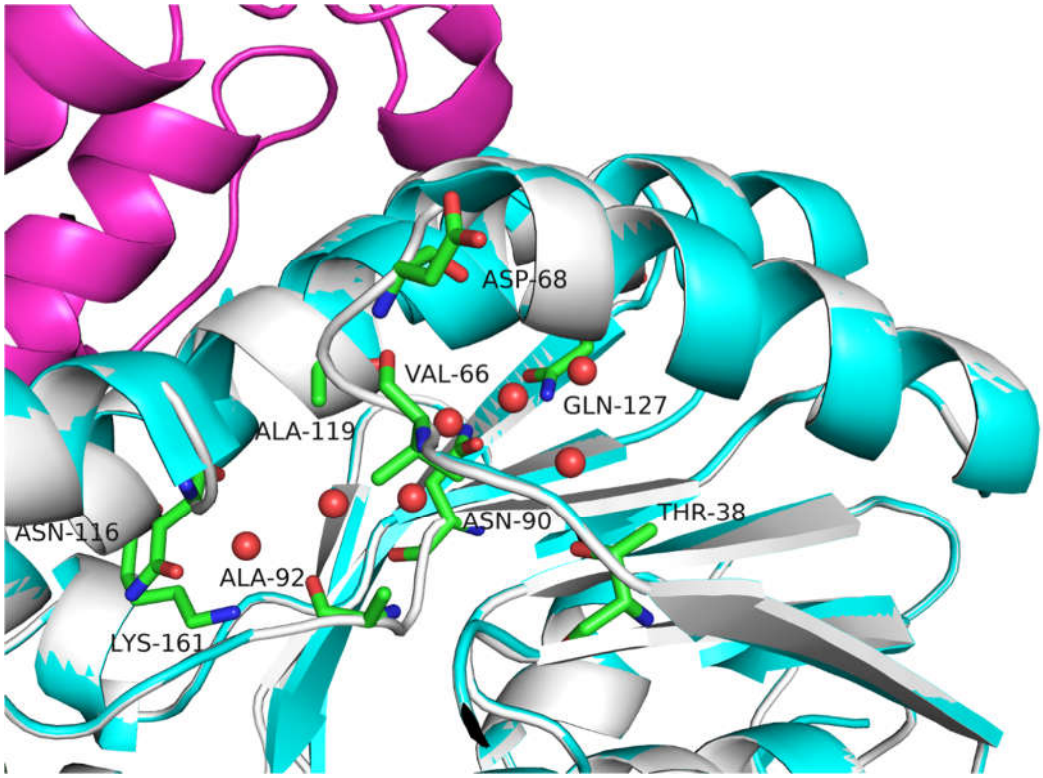


Figure S3.6. Hydrophilic and water-rich channel that shuttle proton from Asn-116 to the solvent bulk (shown in stick and cartoon)