

Structural Studies of MenD – A Crystallographic Endeavor

**A Thesis Submitted to the College of Graduate Studies and
Research in Partial Fulfillment of the Requirements for the Degree
of Master of Science in the Department of Chemistry**

University of Saskatchewan

Saskatoon, Canada

By

Dan Toogood



© Copyright Dan Toogood, October 2008. All rights reserved

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of materials in this thesis in whole or in part should be addressed to:

Head of Department of Chemistry
University of Saskatchewan
110 Science Place
Saskatoon, Saskatchewan
Canada, S7N 5C9

ABSTRACT

The thesis presented here describes the steps that were taken in an attempt to solve the protein structure of MenD via molecular replacement and multiple wavelength anomalous dispersion. The introduction provides background on menaquinone biosynthesis and the role of MenD in this metabolic pathway. Also, a detailed discussion of the DC Family of enzymes, a subgroup of ThDP dependent enzymes, which MenD is a part of, is included.

Utilizing various software packages a 1.9 Å data set was processed and analyzed in an attempt to provide a molecular replacement result. When molecular replacement was deemed incapable of solving the phase problem of the data set, the production of SeMet protein was attempted to allow for MAD phasing.

A homology model of MenD was produced using the program *Modeller* with benzaldehyde lyase as a template. A structure based sequence alignment was done with all DC Family enzymes with structures published. Then a second structure based sequence alignment was done to compare the same set to the *Modeller* model. This was done to gain a deeper understanding of MenD and how it interacts with its cofactors ThDP and Mg²⁺. Furthermore, these results were used to implicate potential active site residues.

ACKNOWLEDGEMENTS

First, I would like to thank my co-supervisor, Dr. David A. R. Sanders, for his guidance, assistance, patience and support during my time as his student. I perhaps cannot begin to understand the magnitude of knowledge he provided me within the field of crystallography during my graduate work.

I wish to also thank my co-supervisor, Dr. David R. J. Palmer, for his expertise and always delightful and energetic conversation in regards to my project. I am grateful to my advisory committee members: Dr. M. Soledade C. Pedras and Dr. Matthew F. Paige for their time and comments during my time here.

Deepest gratitude goes out to the members of the Sanders laboratory group: Karin, Josiah, Sarathy, Eshwari, Sara and various other members. I also owe a great deal of appreciation to the members of Dr. Palmer's lab.

Special regards to Dr. Swarnam Ravindran for her belief in me, advice, and her exceptional support and guidance through the personal challenges I faced during my time as a graduate student.

I would like to thank my family for their support of me in an effort that had never been attempted within my family and perhaps will never be purely understood.

Finally, I wish to thank Mr. Landon Ziffle for without his friendship, accommodation, advice, continuous council and motivation I certainly would not have completed my project, let alone this thesis.

THESIS DEDICATION

To my grandfather,

Lt. Samuel James Toogood

Aug 6th, 1915 - Feb 2nd, 2007

Your Legacy Lives On

Table of Contents

Permission to use.....	i
Abstract.....	ii
Acknowledgements.....	iii
Dedication.....	iv
Table of Contents.....	v
List of Figures.....	viii
List of Tables.....	xiv
List of Abbreviations.....	xv
<u>CHAPTER 1: INTRODUCTION</u>	1
1.1 Menaquinones (Vitamin K ₂).....	1
1.2 Menaquinone Biosynthetic Pathway.....	2
1.3 MenD (SEPHCHC Synthase).....	4
1.4 Thiamine diphosphate dependent enzymes.....	7
1.5 Enzymes of the DC Family.....	17
1.5.1 DC Family Enzymes that Utilize H ⁺ as a Secondary Substrate and Have a Pair of Histidine Residues Required for Catalysis.....	17
1.5.2 DC Family enzymes that utilize the same or similar primary and secondary substrates.....	23
1.5.3 Other DC Family enzymes with Elucidated Structures.....	28
1.5.3.1 Carboxyethyl Arginine Synthase.....	28
1.5.3.2 Pyruvate Oxidase.....	30
1.5.3.3 Oxalyl-CoA Decarboxylase.....	32
1.6 MenD.....	34

<u>Chapter 2: Research Goals and Approach</u>	36
<u>Chapter 3: Materials and Methods</u>	38
3.1 General.....	38
3.2 Software Utilized.....	38
3.3 Cloning Procedures.....	39
3.4 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	40
3.5 Expression Procedures.....	41
3.5.1 Protein Expressed from pMD80L plasmid.....	41
3.5.2 Seleno-methionine Protein Expressed from pMD80L.....	42
3.6 Purification Procedures.....	43
3.6.1 Purification of <i>E. coli</i> MenD From pMD80L Construct.....	43
3.6.2 Purification of SeMet <i>E. coli</i> MenD.....	45
3.7 Dialysis Procedures.....	45
3.8 Concentration Procedures.....	46
3.9 Bradford Assay Analysis to Determine the Extinction Coefficient (ϵ) of MenD	46
3.10 Crystallization Procedures.....	47
3.10.1 Sitting Drop Screening.....	47
3.10.2 Microbatch Screening.....	48
3.10.3 Hanging Drop and Optimization Methods.....	48
3.11 Diffraction Procedures.....	49
<u>Chapter 4: Experimental Results</u>	50
4.1 Computational Work.....	50
4.1.1 MOSFLM and SCALA.....	50

4.1.2 MR BUMP.....	52
4.1.3 AMoRe.....	52
4.1.4 PHASER.....	54
4.1.5 <i>Modeller</i>	58
4.2 Expression of the MenD Gene.....	63
4.2.1 Expression of MenD from pMD14b in BL21.....	63
4.2.2 Selection of a New Vector.....	63
4.2.3 Cloning of MenD gene into pQE80L.....	65
4.2.4 Expression of MenD from pQE80L Construct.....	67
4.2.5 Testing Purified Protein for MenD Activity.....	68
4.2.6 Crystal Trials and diffraction of Native pMD80L protein.....	69
4.2.7 Growth of pMD80L in Minimal Media.....	73
4.2.8 Purification of MenD from pMD80L in Presence of TCEP.....	74
4.2.9 Expression of SeMet Protein from the pMD80L construct.....	75
4.2.10 Crystal Trials of SeMet pMD80L Protein.....	78
4.3 Structure Based Sequence Alignment of ThDP Dependent Enzymes.....	78
4.3.1 Structure based sequence alignment of ThDP Dependent Enzymes using BAL (2AG0) as the Search Model.....	78
4.3.2 Structure Based Sequence Alignment with MenD <i>Modeller</i> Model.....	83
4.3.3 MenD <i>Modeller</i> Model Active Site Residue Analysis.....	91
<u>Chapter 5: Concluding Statements and Summary</u>	93
<u>Chapter 6: References</u>	100

List Of Figures

- Figure 1.1** - Menaquinone molecules MK-2 and MK-4 showing the variable number of isoprenoid residues.....2
- Figure 1.2** - The *men* operon in *E. coli* K12. The six enzymes MenB, MenC, MenD, MenE, MenF and yfbB (MenH) are responsible for the first six steps of the menaquinone pathway. The figure indicates that the MenD gene is approximately 1.7 kilo base pairs. ^(Ecogene Webpage)2
- Figure 1.3** - Menaquinone Biosynthesis Pathway.....4
- Figure 1.4** – Proposed MenD catalytic cycle for the formation of SEPHCHC.....6
- Figure 1.5** – Thiamine diphosphate molecule which contains a pyrimidine ring, a thiazolium ring and a diphosphate group.....7
- Figure 1.6** - First half reactions for: A) generalized decarboxylase reaction B) generalized transferase reaction, note the similarities in the first half reaction of each.....8
- Figure 1.7** - Ylide formation – ThDP is activated for nucleophilic attack upon binding to ThDP-dependent enzymes. An invariant glutamate acts as a proton-acceptor. An ylide is then formed as C2 of the thiazolium ring donates a proton to the pyrimidine ring.....9
- Figure 1.8** - An enamine intermediate is formed following the removal of R₄. The tautomerization allows for both nucleophilic attack by the carbanion intermediate on the secondary substrate and also stabilizes the intermediate in the enamine form.....10
- Figure 1.9** - Transketolase dimer, RCSB protein accession (PDB Code) 1TKB. ThDP is shown as spheres each monomer is shown in blue and green respectively. The overlap of the two monomers creates two equivalent active sites at the dimer interfaces.....11
- Figure 1.10** - Domain organization of ThDP dependent enzymes. Three domains are shown with a varied number of parallel β -sheets, each linked by an α -helix forming a $\alpha\beta$ -type supersecondary fold. The N-terminal PYR domain is on the middle right (blue), the middle domain is shown at the bottom left (green/yellow) and the PP domain is shown in the top left (orange and red). Mg²⁺ is indicated by a salmon colored sphere and ThDP is shown in gray. The image is a monomer subunit of benzoylformate decarboxylase (RCSB Protein Accession: 1BFD).....13
- Figure 1.11** - Catalytic Duality, active site non-equivalence and the 20 Å Tunnel. 1) The electron shuttle through the 20 Å Tunnel between non-equivalent active sites from active site one to active site two when ThDP binds. 2) ThDP ylide is formed within active site one while active

site two is deactivated following the electron shuttle. 3) Reaction with substrate in active site one results in a shuttle of electrons back through the tunnel to active site two. 4) Enamine formation in active site one and ylide formation in active site two.....15

Figure 1.12 – Superimposition of ThDP moieties and GDG-(X₂₄)-NN binding motifs from benzaldehyde lyase (Blue - PDB Code: 2AGO), carboxyethyl arginine synthase (Green – PDB Code: 1UPA) and indolepyruvate decarboxylase (Red – PDB code: 1OVM). The Mg²⁺ ion (shown in magenta) was added manually to indicate its approximate expected location.....16

Figure 1.13 – The catalytic cycle of ScPDC and the general reaction that produces acetaldehyde from pyruvate via ScPDC catalysis are indicated.....19

Figure 1.14 - ScPVD model (RCSB Protein Accession (PDB code): 1PVD) with ThDP shown as orange spheres.....19

Figure 1.15 – The catalytic cycle of ZmPDC and the general reaction that produces acetaldehyde from pyruvate via ZmPDC catalysis are indicated.....20

Figure 1.16 - ZmPDC model (PDB code: 1ZPD) with ThDP shown as orange spheres.....20

Figure 1.17- The catalytic cycle of IPDC and the general reaction that produces indole-3-acetaldehyde from indole-3-pyruvic acid via BFD catalysis are indicated.....21

Figure 1.18 - IPDC model (PDB Code: 1OVM) with ThDP depicted as orange spheres.....21

Figure 1.19 - The catalytic cycle of BFD and the general reaction that produces benzaldehyde from benzylformate via BFD catalysis are indicated.....22

Figure 1.20 – BFD Model (PDB Code: 1BFD). ThDP is depicted in orange as a stick model. The monomer is shown as presented within the protein data bank.....22

Figure 1.21 - The catalytic cycle of AHAS. This enzyme can catalyze two separate reactions to produce 2-acetolactate and 2-aceto-2-hydroxybutyrate.....25

Figure 1.22 - AHAS model (PDB Code: 1JSC). FAD is depicted as spheres and ThDP is depicted in orange as sticks.....25

Figure 1.23 - The catalytic cycle of ALS and the general reaction that indicates the production of 2-acetolactate from two molecules of pyruvate.....26

Figure 1.24 - ALS model (PDB Code: 1OZG) with ThDP depicted in orange as a stick model; note the absence of FAD in the structural representation.....26

Figure 1.25 - BAL catalytic cycle indicating the reversible cleavage and formation of *R*-benzoin to/from two molecules of benzaldehyde.....27

Figure 1.26 – BAL model (PDB Code: 2AG0). ThDP is indicated with orange spheres.....	27
Figure 1.27 - CEAS catalytic cycle indicating the formation of N ² -(2-carboxyethyl)arginine from D-glyceraldehyde and L-arginine.....	29
Figure 1.28 - CEAS model (PDB Code: 1UPA). ThDP in indicated with orange spheres.....	29
Figure 1.29 - POX catalytic cycle indicating the formation of acetylphosphate from pyruvate, oxygen and inorganic phosphate; along with the reduction of FAD.....	31
Figure 1.30 - POX Model (PDB Code: 1POX). FAD is depicted as spheres and ThDP is indicated with orange stick models.....	31
Figure 1.31 - OCDC catalytic cycle indicating the formation of formyl-CoA from oxalyl-CoA.....	33
Figure 1.32 - OCDC Model (PDB Code: 2JIB). ADP and CoA are represented as spheres and ThDP depicted as an orange stick models.....	33
Figure 4.1 - PDB coordinate file produced by PHASER using 2AG0 polyalanine tetramer, Z-score = 10.34, LL-gain (refined) = 79.50.....	58
Figure 4.2 - Ribbon Diagram of monomer model produced by <i>Modeller</i> containing the MenD sequence; viewed in PyMol.....	60
Figure 4.3 - A) generated dimer from the <i>Modeller</i> model B) generated tetramer from the <i>Modeller</i> model.....	61
Figure 4.4 - Superimposition of MenD <i>Modeller</i> model tetramer (blue) and the PDB file produced by PHASER (red) in molecular replacement. Only notable difference is seen in the random loop of the middle domain seen at the top and bottom of the figure in red. LSQMAN reports RMSD = 0.10 Å	62
Figure 4.5 - SDS-PAGE gel showing: A) inadequate production of MenD protein (from pMD14b plasmid) B) adequate production of MenD protein (from pQE80L plasmid). The one noticeable band in figure A is the lysozyme band from the cell digestion step. The molecular weight for MenD is expected to be thought 61.4 KDa.....	64
Figure 4.6 - Map of pQE80L as provided by Qiagen ^(Qiagen 2003)	65
Figure 4.7 - 10% Agarose gel with 1Kb Plus Marker (lane 1, Kb pairs shows beside) and PCR amplified MenD gene following a) single digest with BamHI, b) single digest with KpnI, c) double digest. The experiment shows that the target gene (1.7 Kb pairs) was cut as intended and did not contain internal KpnI or BamHI sites as was expected.....	66

Figure 4.8 - SDS-PAGE Gel: Lane 1 – marker, Lane 2 - 3μL of concentrated and purified protein, Lane 3 – 10μL of concentrated and purified protein, Lane 4 – unpurified soluble fraction, Lane 5 – insoluble pellet protein.....	68
Figure 4.9 - Lactate dehydrogenase assay for monitoring MenD Activity. The spontaneous production of SHCHC from SEPHCHC produces pyruvate which is then converted to lactate by lactate dehydrogenase (LDH). The reaction with LDH results in the oxidation of NADH to NAD ⁺ and thusly causes a loss in absorption at 340nm within the reaction mixture. The rate of reaction is determined by the rate of decrease in absorption at 340nm.....	69
Figure 4.10 - Photograph of crystals within well 88 (H4) in the MPD screen. The largest of the three crystals (bottom left) measures 0.21 mm x 0.07 mm x 0.07 mm.....	72
Figure 4.11 - Photograph of Optimized Crystallization condition (35% MPD, 7.5% PEG4000 and 0.1M HEPES buffer pH 7.5) in a sitting drop well. The crystal with a star next to it measures 0.34 mm x 0.068 mm x 0.068 mm.....	72
Figure 4.12 - Photograph of Hanging Drop Crystallization condition (35% MPD, 7.5% PEG4000 and 0.1M HEPES buffer pH 7.5) A protein film is noticeable in this picture because the photo was taken following exposure to open air to collect crystals for diffraction. Crystal with a red star beside it measures 0.51 mm x 0.11 mm x 0.11 mm.....	73
Figure 4.13 - Expression of MenD from pMD80L within minimal media A) SDS-PAGE gel showing no protein production prior to induction with IPTG. B) SDS-PAGE gel of three identical preparations following IPTG induction, harvesting, sonication and lysis of cells but not yet purified. Note that lane 4 is a cellular pellet showing insoluble protein from all three fractions.....	74
Figure 4.14 - Output chromatogram of purification of protein from pMD80L plasmid; Note the magnitude of the peak A ₂₈₀ in the stepwise gradient elution reaches nearly 0.4. Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.....	76
Figure 4.15 - Output chromatogram of purification of protein from pMD80L plasmid in the presence of 2 mM TCEP; protein sample used was identical to the sample used to produce Figure 4.11. Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.....	76
Figure 4.16 - Output chromatogram of purification of SeMet protein from pMD80L plasmid in the presence of 2 mM TCEP Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.....	77
Figure 4.17 - Structure based sequence alignment of the DC family with benzaldehyde lyase as the structural search model.....	81-82

- Figure 4.18** - Structure based sequence alignment of the DC family with the MenD *Modeller* model as the search model. It is important to note that the sequence number listed at the top of each segment indicates the amino acid number minus one. This is due to the lack of methionine at the beginning of the sequence. For example the invariant Glu55 of MenD would appear to be numbered 54 in this figure.....84-85
- Figure 4.19** - Analysis of the 1POX model to ensure Glu59 is the first residue in the alpha helix as is the tendency for the invariant glutamate residue in ThDP-dependent enzymes.....86
- Figure 4.20** - Location of Glu55 within the MenD *Modeller* Model. The distance from the carboxyl group of glutamate to the N1' atom of the pyrimidine is 4.1 Å (measured with Coot) which is slightly out of range for hydrogen bonding. The orientation is as was expected though.....86
- Figure 4.21** - The 18.83 Å tunnel between the two active sites within one of the dimer pairs of the MenD *Modeller* model. The charged residues between the pair of active sites may form a hydrogen bonding bridge that explains the non-equivalence of active sites often seen in ThDP-dependent enzymes (Figure 1.11).....87
- Figure 4.22** - Location of Leu443 within the MenD *Modeller* Model; Note the orientation that runs along the edge of the ThDP binding pocket.....89
- Figure 4.23** - Location of Ile418 within the MenD *Modeller* Model; Note the location is not in close proximity to the cofactor ThDP.....89
- Figure 4.24** - Location of Asp442 within the MenD *Modeller* Model; Note the orientation that runs directly adjacent to the diphosphate and can be assumed to be in close proximity to the Mg²⁺ cofactor. Also note the presence of the remainder of the GDL-X₂₄-N binding motif within this figure and how all of these residues are in direct proximity to the diphosphate of ThDP.....90
- Figure 4.25** - The MenD active site residues according to the MenD *Modeller* Model; Residues thought to be capable of participating in catalysis/binding of substrate include (but are not limited to) Asn117, Gln118, Ser32, Arg33 and Leu478.....92
- Figure 5.1** – The structural model of MenD (gold) superimposed over the *Modeller* model (blue). ThDP is shown as a stick model in pink. Noticeable differences are seen between the two models, specifically in the middle domain where the six parallel β-sheets do not overlap.....95
- Figure 5.2** – Location of both Ile418 and Leu443 in the MenD model. This indicates that the V-conformation is induced by Ile418 and not Leu443.....96

Figure 5.3 – The active site of MenD. Arg33, Arg107, Lys292 and Arg295 aid in the binding of MenD’s acidic substrates, isochorismate and 2-ketoglutarate. Hydrophobic contacts are made with the hydrophobic ring of isochorismate by Ile474, Phe475, and Leu478. Note that the active site is formed from a dimer, monomer A is shown in green and monomer B is shown in light blue.....97

Figure 5.4 – ThDP-binding motif GDL-(X₂₄)-NN within the MenD structure.....98

Figure 5.5 – The absence of a catalytic duality tunnel in the structural model of MenD. The residues Asp54, Glu55, Arg56 and Tyr447 block access between the two active sites.....99

List of Tables

Table 3.1 - SDS-PAGE gel preparation recipe for one gel (*APS must be made fresh).....40

Table 4.1 - Output data from SCALA to determine the resolution limit. The first column of each table indicates the series of values of increasing resolution. The second column indicates a value of R_{merge} ; a value of 0.400 was utilized to select resolution limit with R_{merge} . The third column indicates the signal to noise ratio (I/σ); a value of 2.0 was utilized to select resolution limit with I/σ . Table A shows the holoenzyme data; Table B shows the apoenzyme data.....51

Table 4.2 – Output results from AMoRe. The first column indicates the protein data bank model utilized for analysis. The second column provides the sequence identity of the protein in comparison to the MenD protein sequence. CC (2nd) refers to the best and second best result produced for correlation coefficient as produced by AMoRe. R_f (2nd) refers to the best and second best result for R_f values as produced by AMoRe. The PDB form indicates whether the coordinate file was for a dimer or tetramer. The following two columns are for the polyalanine models. The final column indicates the protein that is being analyzed.....55

Table 4.3 - Output results from PHASER. The first two columns of data indicate the output Z-score and the value of log-likelihood gain following the rotation function. The next two columns displays the output Z-score and the value of log-likelihood gain following the translation function. The LL-gain refined indicates the log-likelihood gain following anisotropic corrections.....57

Table 4.4 - Crystallization conditions that produced crystals at [protein] of 10mg/mL and diffraction results from the in-house DX8 Proteum diffractometer.....71

List of Abbreviations

ADP – Adenosine diphosphate

AEBSF - 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride

AHAS – Acetohydroxy acid synthase

ALS – Acetolactate synthase

AMoRe – A program for molecular replacement

APS – Ammonium persulphate

BAL – Benzaldehyde lyase

BFD – Benzoylformate decarboxylase

CEAS – Carboxyethyl arginine synthase

ComPAS – Complementary screen of polymers, alcohols and salts

DC Family – Decarboxylase family

DHNA-CoA - 1,4-Dihydroxyl-2-naphthoyl-CoA

DMK-n – Demethylated menaquinone where n is the number of repeat isoprenoid residues

DTT - Dithiothreitol

EDTA - Ethylenediaminetetraacetic acid

FAD – Flavin adenine dinucleotide

HPLC – High performance liquid chromatography

IPDC – Indole pyruvate decarboxylase

JCSG+ - Crystal screen developed by the Joint Center for Structural Genomics

LB – Lauri Bertani broth

MAD – Multiple wavelength anomalous dispersion

MK-n – Menaquinone where n is the number of repeat isoprenoid residues

MPD – 2-Methyl-2,4-pentanediol

MR – Molecular replacement

MRBUMP – Molecular replacement by using multiple proteins

NADH – Nicotinamide adenine dinucleotide

NMR – Nuclear magnetic resonance

OCDC – Oxalyl-CoA decarboxylase

OD₆₀₀ – Optical density at 600nm

OSB - *o*-Succinylbenzoate

OSB-CoA - *o*-Succinylbenzoyl-CoA

PACT – Screen containing a pH, anions and cations template

PCR – Polymerase chain reaction

PDB – Protein data bank

PDC – Pyruvate decarboxylase

PEG – Polyethylene glycol

POX – Pyruvate oxidase

PP – Diphosphate binding domain

Pyr – Pyrimidine binding domain

RMSD – Root mean squared deviation

ScPDC – PDC from *Saccharomyces cerevisiae*

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SeMet – Selenomethionine

SEPHCHC - 2-Succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic acid

SHCHC - (1*R*,6*R*)-2-Succinyl-6-hydroxyl-2,4-cyclohexadiene-carboxylate

TCEP – Tris(2-carboxyethyl) phosphine hydrochloride

ThDP – Thiamine diphosphate

UV/Vis – Ultraviolet/visible spectrum

VKD – Vitamin K dependent enzyme

ZmPDC – PDC from *Zymomonas mobilis*

CHAPTER 1: INTRODUCTION

1.1 Menaquinones (Vitamin K₂)

Menaquinones (Vitamin K₂) are a group of compounds that contain a common naphthoquinone ring structure with a variable number of isoprenoid residues attached to the 3-position. They are generally referred to as MK-n, where n is the number of isoprenoid residues attached (Figure 1.1). Menaquinones are produced by certain bacteria, including *Mycobacterium tuberculosis* and *Bacillus subtilis*, for utilization as an electron shuttle within the electron transport chain (Bishop 1962; Truglio 2003).

Humans cannot synthesize Vitamin K₂ and as such it must be obtained in our diet or from bacteria present within our intestine, such as *Escherichia coli* (Bentley 1982). Vitamin K₂ plays an essential role as a cofactor in the carboxylation of glutamate from a variety of proteins known as Vitamin K dependent enzymes (VKD's). VKD's are now known to be present in almost all human tissues and are also required within bacteria. Among the many roles VKD's serve, they have been shown to be required for the processes of bone mineralization (Weber 2001), blood coagulation (Vermeer 2001), apoptosis (Cheng 2003), signal transduction (Kulman 1997; Kulman 2001), growth control (Ni 1998), phagocytosis (Hall 2001; Hall 2002) and chemotaxis (Fridell 1998). It has also been proven that the medicinal provision of menatetrenone, MK-4, improves bone mineral density in osteoporosis patients (Shiraki 2000).

1.2 Menaquinone Biosynthetic Pathway

The biosynthesis of menaquinones within *E. coli* has been studied thoroughly (Meganathan 2001). Synthesis of menaquinones is performed by various enzymes encoded by the menaquinone operon, or as it is sometimes referred to, the *men* gene cluster (Figure 1.2). Within this gene cluster are six open reading frames which encode six enzymes named MenB through MenF and *yfbB* respectively.

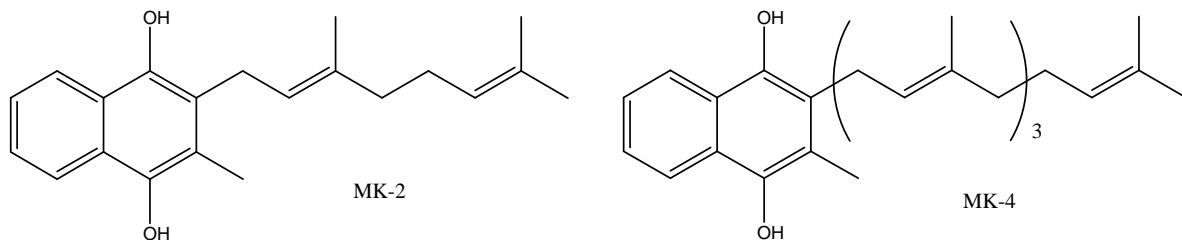


Figure 1.1 - Menaquinone chemical structures MK-2 and MK-4 showing the variable number of isoprenoid residues

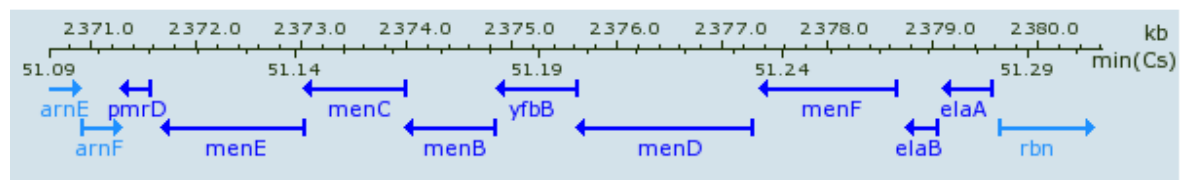


Figure 1.2 - The *men* operon in *E. coli* K12. The six enzymes MenB, MenC, MenD, MenE, MenF and *yfbB* (MenH) are responsible for the first six steps of the menaquinone pathway. The figure indicates that the *MenD* gene is approximately 1.7 kilo base pairs (Ecogene Webpage).

These six enzymes encompass a large portion of the molecular machinery required for the conversion of chorismate to menaquinones (Figure 1.3). The first reaction is the isomerization of the shikimate pathway product chorismate to isochorismate via the enzyme MenF. Although other enzymes produce isochorismate, MenF is thought to be the lone menaquinone pathway-specific isochorismate synthase (Daruwala 1996). It was understood that the subsequent reaction is the conversion of the MenF product isochorismate to (1*R*,6*R*)-2-

succinyl-6-hydroxyl-2,4-cyclohexadiene-carboxylate (SHCHC) via MenD (SHCHC synthase), acting as both a decarboxylase and a lyase (Palaniappan 1992). However, it has been discovered recently that MenD produces 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic acid (SEPHCHC) from isochorismate rather than SHCHC (Jiang 2007). The formation of SHCHC from SEPHCHC is completed by MenH, which is also sometimes referred to as yfbB (Jiang 2008). The subsequent reaction in the pathway involves the conversion of SHCHC to *o*-succinylbenzoate (OSB) via a dehydration reaction performed by MenC (*o*-succinylbenzoate synthase) (Sharma 1993). OSB is then linked to coenzyme A via MenE (*o*-succinylbenzoate-CoA ligase) to produce *o*-succinylbenzoyl-CoA (OSB-CoA) (Sharma 1996) and then MenB (dihydroxynaphthoic acid synthetase) performs a ring closing reaction to produce the naphthoquinone ring structure seen in 1,4-dihydroxy-2-naphthoyl-CoA (DHNA-CoA) (Sharma 1992).

As is noted in Figure 1.3, there is still an unknown enzymatic reaction that takes place within the pathway. It was previously believed that MenH converted DHNA-CoA to 1,4-dihydroxyl-2-naphthoate (DHNA) (Meganathan 2001). However, this has been proven to not be the case and further research is required for a complete understanding of the menaquinone pathway (Jiang 2008).

The differentiation of the various MK-n molecules is a result of various enzymes that act on DHNA. The products vary by the number of isoprenoid repeat units that are added to DHNA. The most commonly discussed, is the formation of MK-8 via MenA and MenG. The product of MenA is a demethylated form of MK-8 referred to as DMK-8 (Suvarna 1998) and finally MenG completes the menaquinone pathway by converting DMK-8 to MK-8 (KoikeTakeshita 1997).

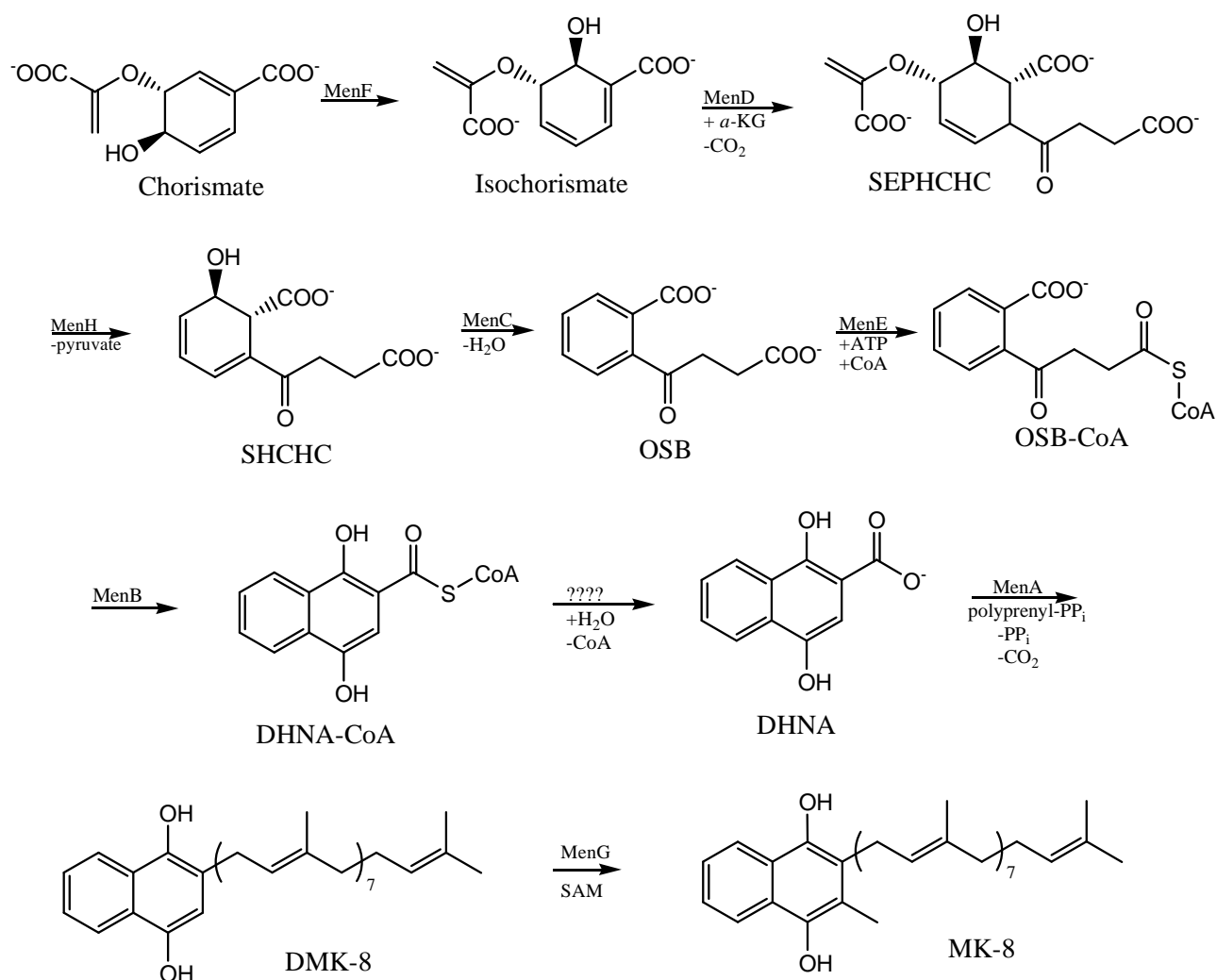


Figure 1.3 - Menaquinone Biosynthetic Pathway

1.3 MenD (SEPHCHC Synthase)

Until the early 1980's it was believed that chorismate was directly converted to OSB via an enzymatic reaction involving 2-ketoglutarate (Meganathan 1981). However, further studies showed that isochorismate was involved in this reaction not chorismate, and also, a separate intermediate was formed via an enzymatic reaction (Bentley 1983; Meganathan 1983). This

intermediate was SHCHC and the enzyme was the product of the *menD* gene (Emmons 1985). The MenD reaction was thought to involve isochorismate and 2-ketoglutarate in the presence of thiamine diphosphate (ThDP) and a divalent cation such as Mg^{2+} to produce SHCHC, pyruvate and CO_2 . The rate of this reaction was monitored via a reduction in absorbance at 340 nm via a coupled assay with lactate dehydrogenase and NADH which reacts with the pyruvate product of the reaction (Bhasin 2003).

Recently, it has been proposed that SHCHC is in fact not the product of the MenD reaction (Jiang 2007). This conclusion was made when it was uncovered that the consumption of isochorismate was not proportional to the formation of SHCHC. In acidic condition MenD consumes the isochorismate substrate but no SHCHC is formed, however, a new compound which was thought to be an intermediate, SEPHCHC, is formed (Figure 1.4). Under basic conditions (pH 7.4 – pH 10) the formation of SHCHC occurs at a slow rate (faster at higher pH values), even in the absence of MenD enzyme. The incorrect assignment of SHCHC as the product of MenD is most likely due to the spectrophotometric properties of SEPHCHC that make it undetectable by UV/Vis spectroscopy. It was not until separation and NMR analysis (Jiang 2007) that the structure of this intermediate was confirmed. The rate of the enzymatic reaction was determined on the basis of isochorismate consumption via HPLC or UV/Vis spectrometry, both utilizing absorbance of photons emitted at 278 nm, the peak absorbance of isochorismate.

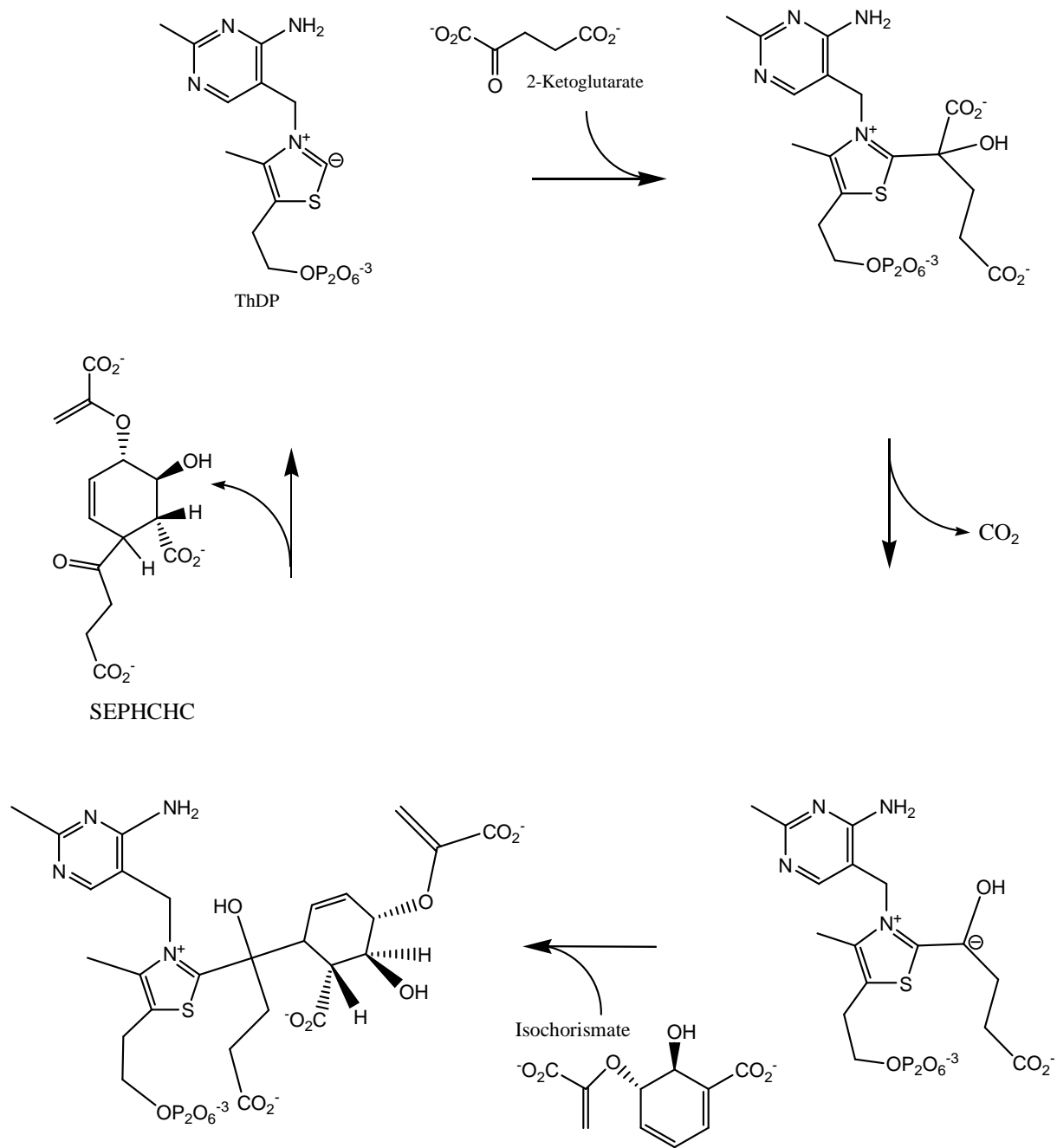


Figure 1.4 – Proposed MenD catalytic cycle for the formation of SEPHCHC.

1.4 Thiamine diphosphate dependent enzymes

Enzymes that use thiamine diphosphate (Figure 1.5) as a cofactor participate in many biosynthetic pathways. Among these many pathways are the key energy-providing pathways, such as the pentose phosphate pathway and the tricarboxylic acid cycle (Frank 2007).

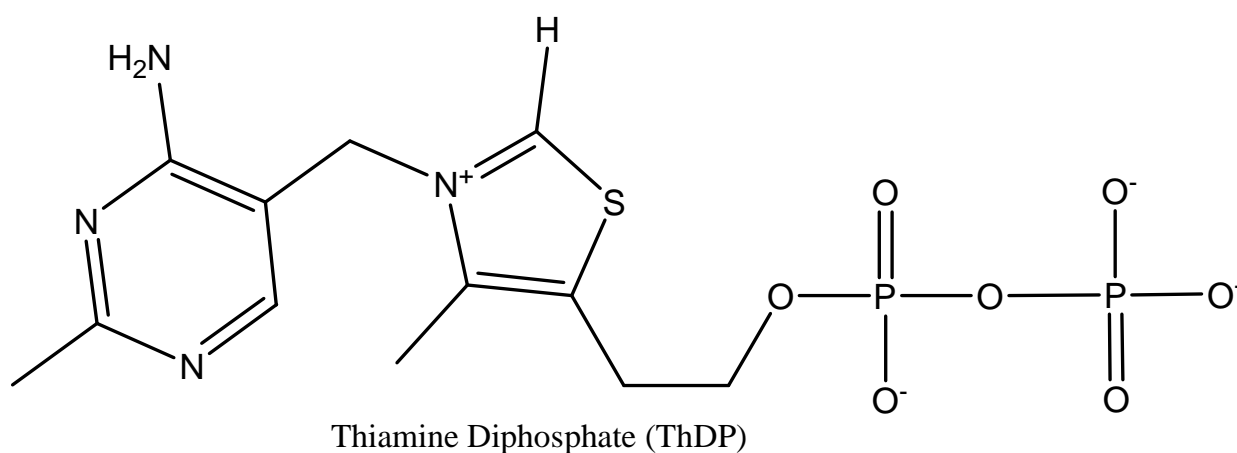


Figure 1.5 – Thiamine diphosphate chemical structure which contains a pyrimidine ring, a thiazolium ring and a diphosphate group.

Some generalizations can be made in regards to the reactions that are occurring within ThDP-dependent enzymes. For instance, the enzymes can be classified into two general categories, those that perform decarboxylation reactions and those that perform non-decarboxylating transferase-type reactions. It is notable that the substrates of the decarboxylation reactions are often 2-ketoacids and the transferase reactions often involve aldehydes (Frank 2007).

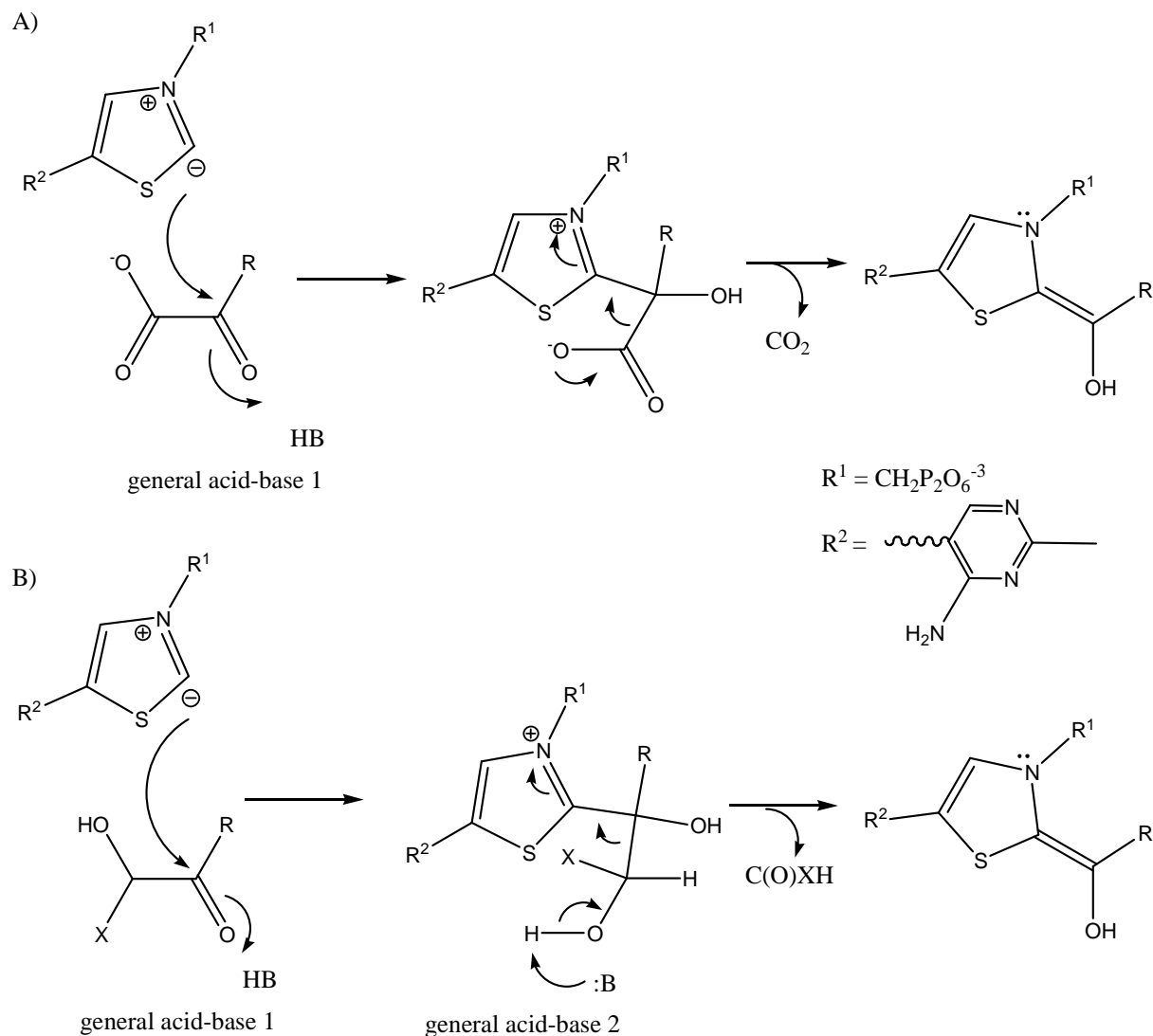


Figure 1.6 - First half reactions for: A) generalized decarboxylase reaction B) generalized transferase reaction, note the similarities in the first half reaction of each

Both the decarboxylase and transferase reactions (Figure 1.6) are initiated by the ThDP moiety. Upon binding of ThDP to the active site, an ylide is formed (Figure 1.7). Ylide formation allows for a nucleophilic attack from the C2 position of the thiazolium ring of ThDP on the substrate, enabling C-C bond lysis adjacent to a carbonyl in the substrate. The substrate is then covalently bound to the ThDP and forms an enamine intermediate with which a second

substrate can react (Figure 1.8). The nature of the bound substrate and the secondary substrate allows for the formation of a diverse library of products from ThDP dependent enzymes.

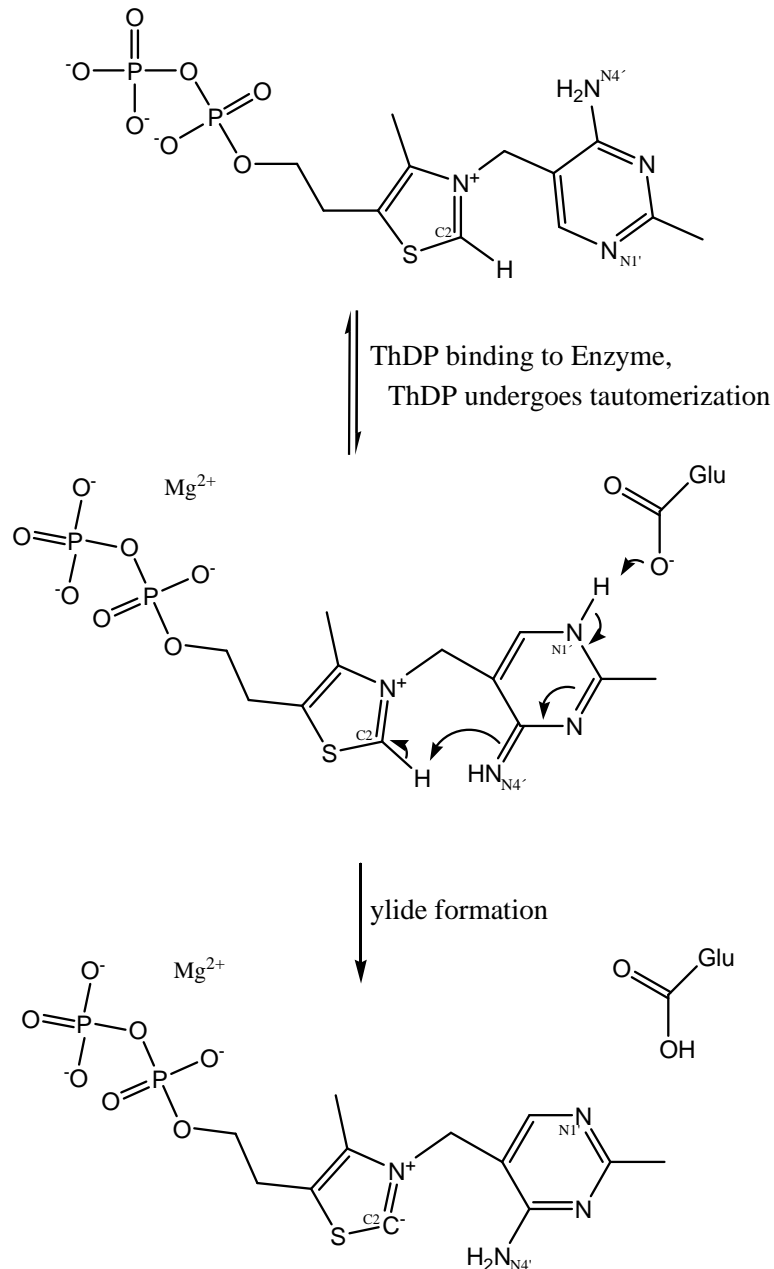


Figure 1.7 - Ylide formation – ThDP is activated for nucleophilic attack upon binding to ThDP-dependent enzymes. An invariant glutamate acts as a proton-acceptor. An ylide is then formed as C2 of the thiazolium ring donates a proton to the pyrimidine ring.

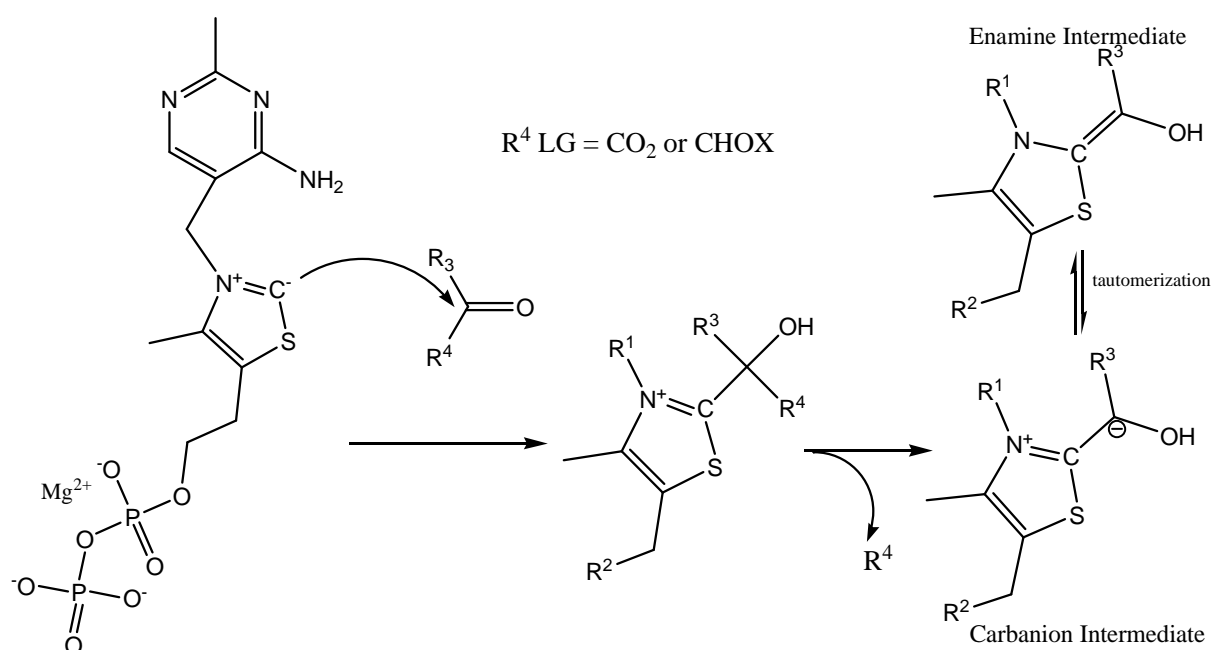


Figure 1.8 - An enamine intermediate is formed following the removal of R^4 . The tautomerization allows for both nucleophilic attack by the carbanion intermediate on the secondary substrate and also stabilizes the intermediate in the enamine form.

The variation in substrates is due to differences seen with the active site residues of the enzymes within the family. However, the conserved binding of a divalent cation, the ThDP moiety and the subsequent formation of the enamine can be viewed as evidence of evolution from a common ancestral enzyme (Frank 2007).

Transketolase (TK, Figure 1.9) (Lindqvist 1994) was the first ThDP dependent enzyme with a published crystal structure. Since this finding, the structural models of many ThDP dependent enzymes have been determined via crystallographic methods. These structural models, along with structure-based sequence alignments of known structures and also sequence alignments of unknown structures, has allowed for the classification of ThDP dependent

enzymes into five separate families (Duggleby 2006). The first family is the transketolase family, which comprises enzymes such as dihydroxyacetone synthase and transketolase itself. Another is the oxidoreductase family, which includes enzymes such as pyruvate:ferredoxin oxidoreductase. Two of the other families are the 2-ketoacid dehydrogenases (K1 and K2). The K1 and K2 enzymes are multienzyme complexes such as the pyruvate dehydrogenase complex and the branched-chain 2-ketoacid dehydrogenase E1.

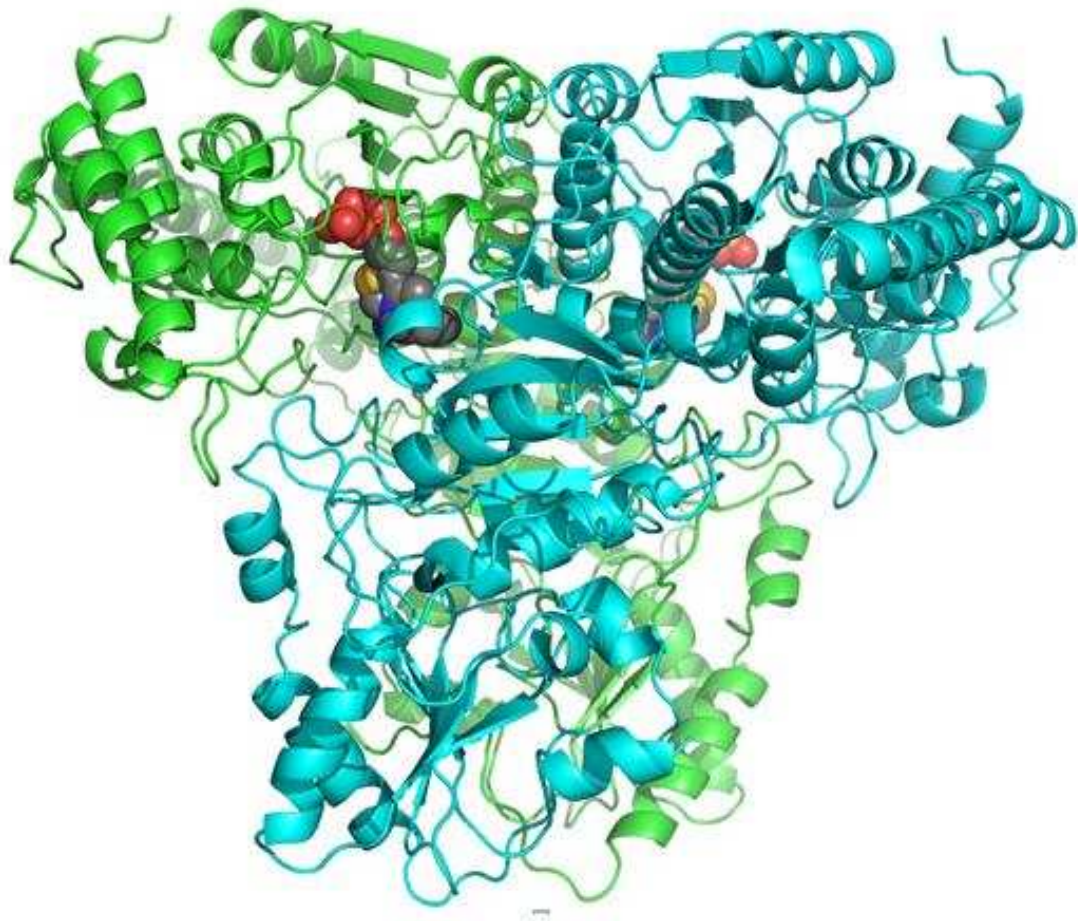


Figure 1.9 - Transketolase dimer, RCSB protein accession (PDB Code: 1TKB). ThDP is shown as spheres each monomer is shown in blue and green respectively. The overlap of the two monomers creates two equivalent active sites at the dimer interfaces.

The final family designation is the decarboxylase family (DC family) where the common reaction includes the decarboxylation of a 2-ketoacid. The enzymes in this family that have had X-ray crystal structures elucidated include benzaldehyde lyase (BAL) (Mosbacher 2005), acetolactate synthase (ALS) (Pang 2004), acetohydroxy acid synthase (AHAS) (Pang 2002), indole pyruvate decarboxylase (IPDC) (Schutz 2003), benzoylformate decarboxylase (BFD) (Hasson 1998), carboxyethyl arginine synthase (CEAS) (Caines 2004), pyruvate oxidase (POX) (Muller 1994), oxalyl-CoA decarboxylase (OCDC) (Berthold 2007), and pyruvate decarboxylase (PDC) from both *S. cerevisiae* (Arjunan 1996) and *Z. mobilis* (Dobritzsch 1998). MenD is also a member of this ThDP-dependent decarboxylase family of enzymes (Duggleby 2006) and as such this family will be discussed in detail.

Within the DC family a common domain arrangement is seen in each monomeric subunit of the enzyme (Figure 1.10). All of the monomer subunits of these enzymes contain three domains that are similar in size. The three domains can be referred to as the N-terminal pyrimidine binding domain (Pyr domain), the middle domain, and the C-terminal diphosphate-binding domain (PP domain). Each of the domains contains a core composed of five or six parallel β -sheets, with each β -sheet linked by an α -helix forming a α/β -type supersecondary fold that resembles a Rossmann fold (Muller 1993). Despite the similarities seen with the three domains with regard to fold, only the Pyr and PP domains are involved in catalysis. All enzymes within the DC family exist as a dimer or a tetramer (dimer of dimers), with two or four active sites, respectively. The Pyr domain from one monomer and the PP domain of the second monomer come together to form one of the active sites on the enzyme. The active site cleft

formed by this association differs from enzyme to enzyme altering the accessibility of the bulk solvent, ThDP, and substrates to the active site.

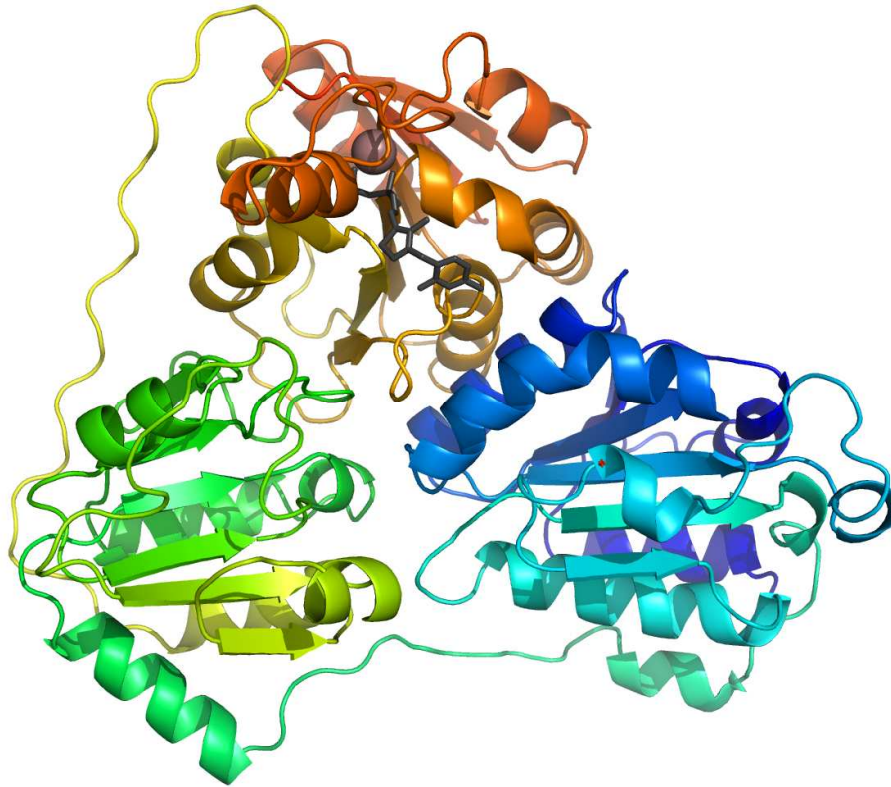


Figure 1.10 - Domain organization of ThDP dependent enzymes. Three domains are shown with a varied number of parallel β -sheets, each linked by an α -helix forming a α/β -type supersecondary fold. The N-terminal PYR domain is on the middle right (blue), the middle domain is shown at the bottom left (green/yellow) and the PP domain is shown in the top left (orange and red). Mg^{2+} is indicated by a salmon colored sphere and ThDP is shown in gray. The image is a monomer subunit of benzoylformate decarboxylase (RCSB Protein Accession: 1BFD)

It has recently been discussed that ThDP dependent enzymes function as a dimer of non-equivalent active sites (Frank 2007). This was first thought to be a possibility when non-equivalence of the active sites and the nature of the substrate and cofactor within it were found to be different in the crystal structure of transketolase (Lindqvist 1994). This non-equivalence is accounted for by the belief that a proton shuttle exists between the two aminopyrimidine rings of

ThDP in the two active sites. The distance has been approximated to be nearly 20 Å between the two ThDP molecules and there exists a cavity filled with hydrogen bonded water molecules and acidic aspartate and glutamate residues (Figure 1.11). This network of hydrogen bonds allows for communication, or a bridge, between the two cofactors.

Both the Pyr and the PP domain are involved in the binding of the cofactors and the catalytic reaction. The Pyr domain provides a hydrophobic cleft for binding the pyrimidine ring of ThDP. The structural and electronic environment of the enzyme active site induces the V conformation of ThDP, placing the N4' atom of the pyrimidine moiety proximal to the C2 atom of the thiazolium ring (Frank 2007). There also exists an invariant glutamate residue within the Pyr domain that interacts with N1' atom of pyrimidine ring of ThDP. This glutamate residue causes the induction of the 1',4'-imino tautomer of the ThDP moiety. When ThDP is in this tautomeric conformation the lone pair of electrons on the imino group of the aminopyrimidine ring is adjacent to the C2 atom of the thiazolium ring. The result of this conformation change is a proton transfer that forms an ylide which can then activate the enzyme substrate via nucleophilic attack (Figure 1.7).

The PP domain is primarily involved in the binding of the diphosphate portion of the ThDP molecule and Mg^{2+} . This interaction primarily relies on a structural motif / sequence fingerprint GDG-(X)₂₄-NN and the resultant coordination between the interactions with Mg^{2+} and the diphosphate moiety within the active site. The interactions made by the residues within this motif and Mg^{2+} involve the aspartate side chain oxygen atoms, the main chain carbonyl of the first glycine residue, the side chain of the second asparagine, as well as coordination with a water molecule (Figure 1.12).

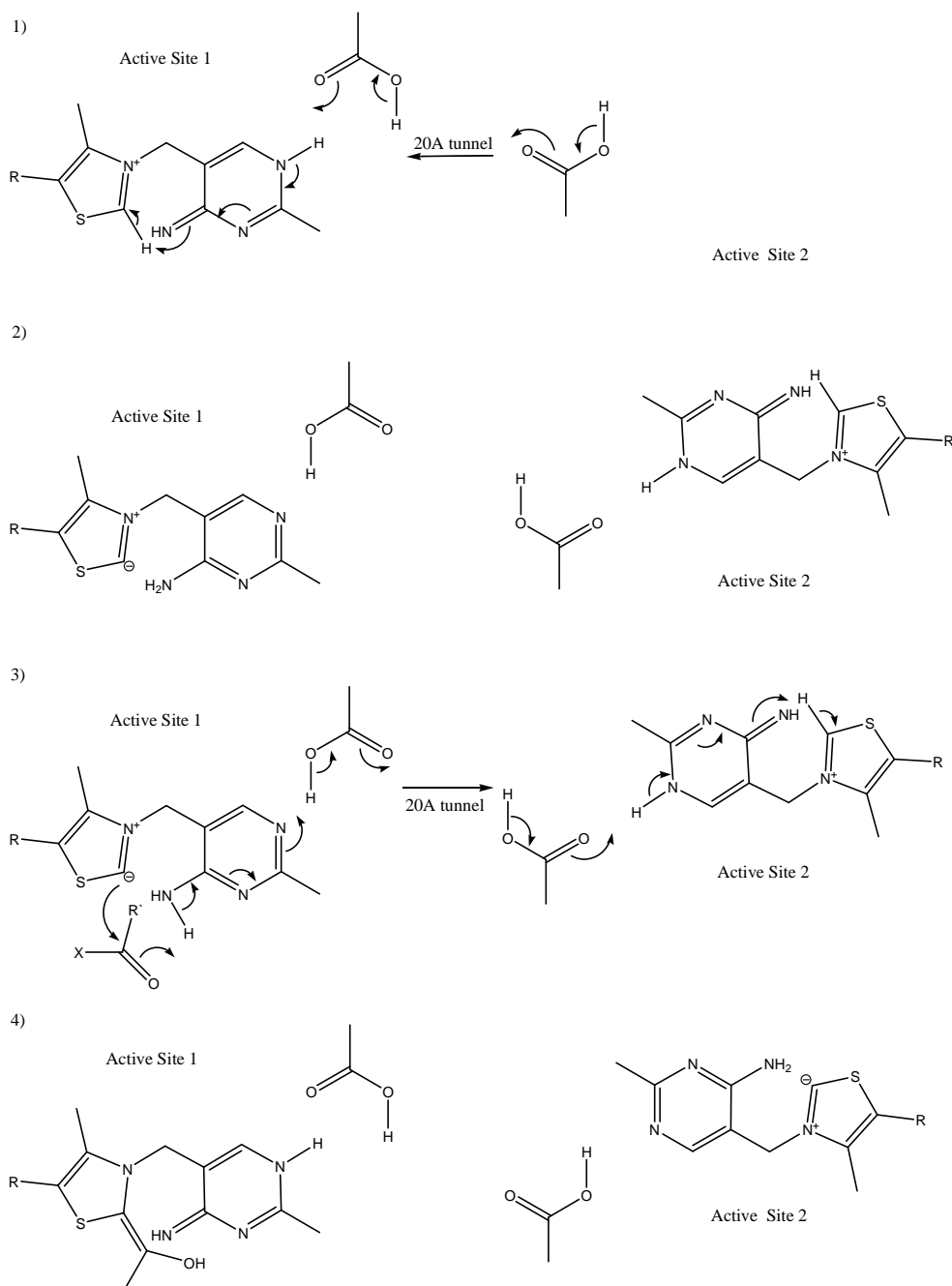


Figure 1.11 - Catalytic Duality, active site non-equivalence and the 20 Å Tunnel. 1) The electron shuttle through the 20 Å Tunnel between non-equivalent active sites from active site one to active site two when ThDP binds. 2) ThDP ylide is formed within active site one while active site two is deactivated following the electron shuttle. 3) Reaction with substrate in active site one results in a shuttle of electrons back through the tunnel to active site two. 4) Enamine formation in active site one and ylide formation in active site two

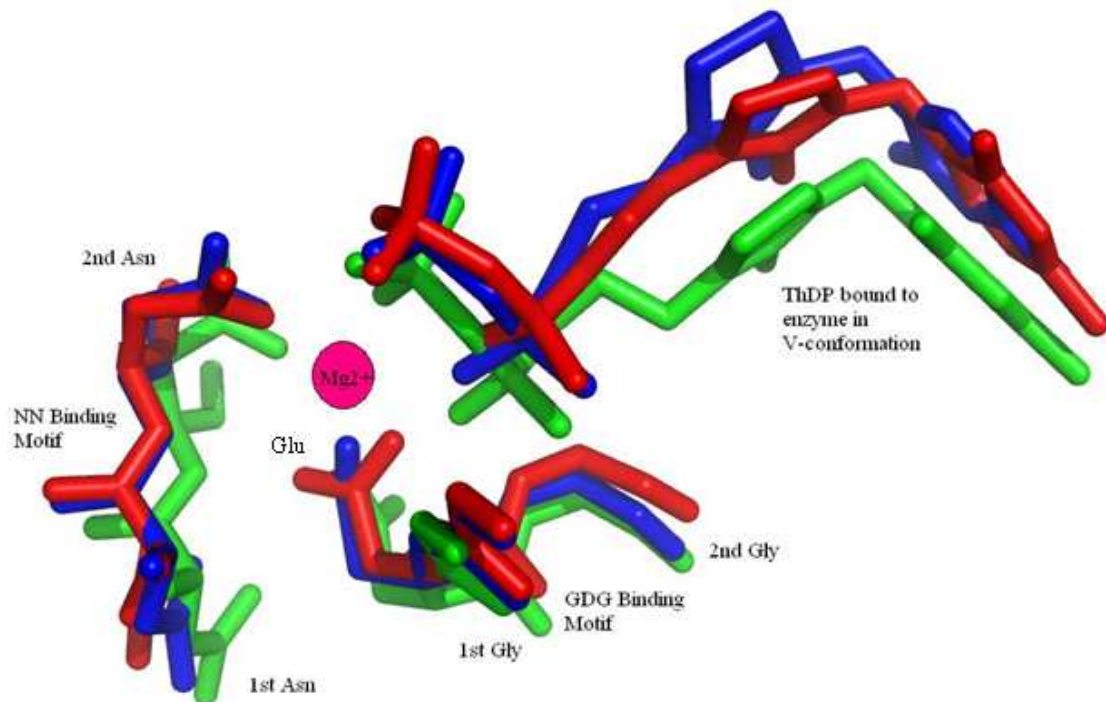


Figure 1.12 – Superimposition of ThDP moieties and GDG-(X₂₄)-NN binding motifs from benzaldehyde lyase (Blue - PDB Code: 2AGO), carboxyethyl arginine synthase (Green – PDB Code: 1UPA) and indolepyruvate decarboxylase (Red – PDB code: 1OVM). The Mg²⁺ ion (shown in magenta) was added manually to indicate its approximate expected location.

As stated earlier, the middle domain appears to have no involvement in the catalytic reaction associated with the enzyme. It also has very little involvement in the dimer interface interactions of the enzyme and only minimal involvement in dimer-dimer interactions seen in tetrameric crystal structures that have been determined. It is therefore thought that this domain exists as a relic of evolution which was required for nucleotide binding (Duggleby 2006). Within POX and AHAS the middle domain is arranged such that flavin adenine dinucleotide (FAD) can be accommodated (Pang 2002, Muller 1993). Within the enzyme OCDC the middle domain allows for the binding of the large oxalyl-coenzyme-A substrate (Berthold 2007).

1.5 Enzymes of the DC Family

1.5.1 DC Family Enzymes that Utilize H⁺ as a Secondary Substrate and Have a Pair of Histidine Residues Required for Catalysis

Four of the ten DC family enzymes with structures elucidated share very common catalytic features. These enzymes are pyruvate decarboxylase from both *Z. mobilis* (*ZmPDC*) (Dobritsch 1998) and *S. cerevisiae* (*ScPDC*) (Arjunan 1996), indolepyruvate decarboxylase (*IPDC*) (Schutz 2003) and benzoylformate decarboxylase (*BFD*) (Hasson 1998).

Within this subgroup of the DC family, the primary substrates that undergo nucleophilic attack from the ThDP ylide are all 2-ketoacids. Also, each of the four enzymes catalyzes non-oxidative decarboxylation reactions of these 2-ketoacid substrates (*ScPDC* Figure 1.13, *ZmPDC* Figure 1.15, *IPDC* Figure 1.17 and *BFD* Figure 1.19). Subsequently, the product is formed with the addition of a proton to the decarboxylated intermediates. The reaction catalysis is believed to be primarily aided by the presence of a pair of histidine residues within the active site (Schutz 2003). Mutagenesis studies of these residues in *ScPDC* (Liu 2001) and *ZmPDC* (Schenk 1997) indicated that these residues were essential for catalysis. The histidine pair for each of the enzymes is as follows: *ZmPDC* (His113, His 114), *ScPDC* (His114, His115), *IPDC* (His115, His116), and *BFD* (His70, His281). It is apparent that the histidine pair in *BFD* would not be indicated with a sequence alignment with the other enzymes. However, the histidines within *BFD* are positionally conserved within the active site (Hasson 1998).

A further similarity between these enzymes is seen with the binding of the common ThDP-dependent cofactors Mg^{2+} and ThDP. The GDG-X₂₄-NN cofactor binding motif is completely conserved in all four enzymes.

The first of these four structures to be elucidated was *Sc*PDC (Figure 1.14). The structure was determined with single wavelength isomorphous replacement (SIR) using *S. uvarum* PDC crystals. The structure of *Zm*PDC (Figure 1.16) was then determined by utilizing the *Sc*PDC model for molecular replacement. Similarly, the *Zm*PDC model was then used to determine the structure of IPDC (Figure 1.18) via molecular replacement. The structure of BFD (Figure 1.20) was determined with anomalous diffraction techniques utilizing mercury soaked crystals.

There are strong indications of evolutionary divergence from a common ancestor when comparing the two PDC's to IPDC. The PDC's share a similar function and also a high sequence identity (31%) with IPDC. However, the most apparent difference that indicates this divergence is the presence of a hydrophobic pocket within the active site of IPDC, which is not present in *Zm*PDC (Schutz 2003). The active site cleft of IPDC is lined with hydrophobic residues such as alanine and leucine, however, in *Zm*PDC the enzyme active site is filled with large bulky hydrophobic side chains, such as tryptophan and tyrosine. These large residues prevent the binding of the larger indolepyruvate by reducing the active site volume of *Zm*PDC. The alteration of Trp392 within *Zm*PDC to Ala387 within IPDC, spatially allows for the binding of the larger indolepyruvate substrate to the active site of IPDC.

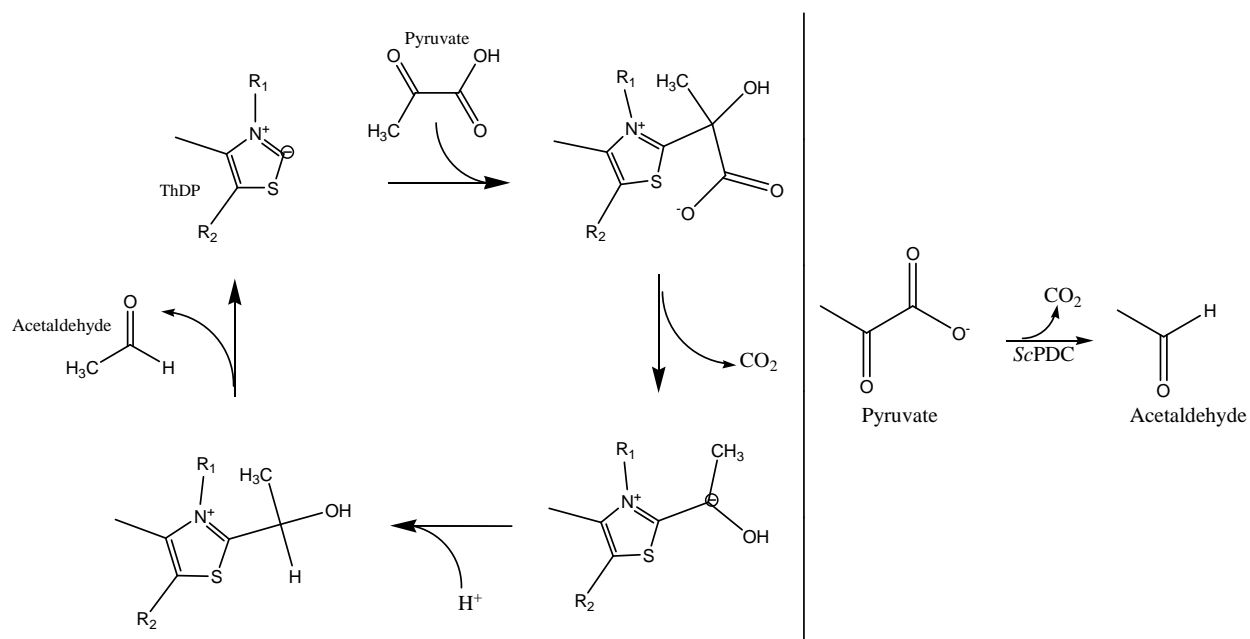


Figure 1.13 – The catalytic cycle of ScPDC and the general reaction that produces acetaldehyde from pyruvate via ScPDC catalysis are indicated.

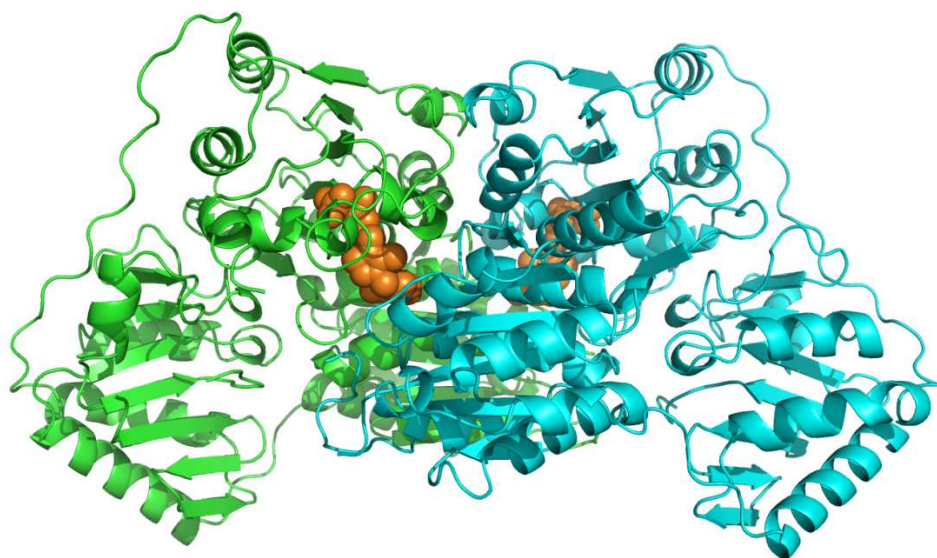


Figure 1.14 - ScPVD model (RCSB Protein Accession (PDB code): 1PVD) with ThDP shown as orange spheres.

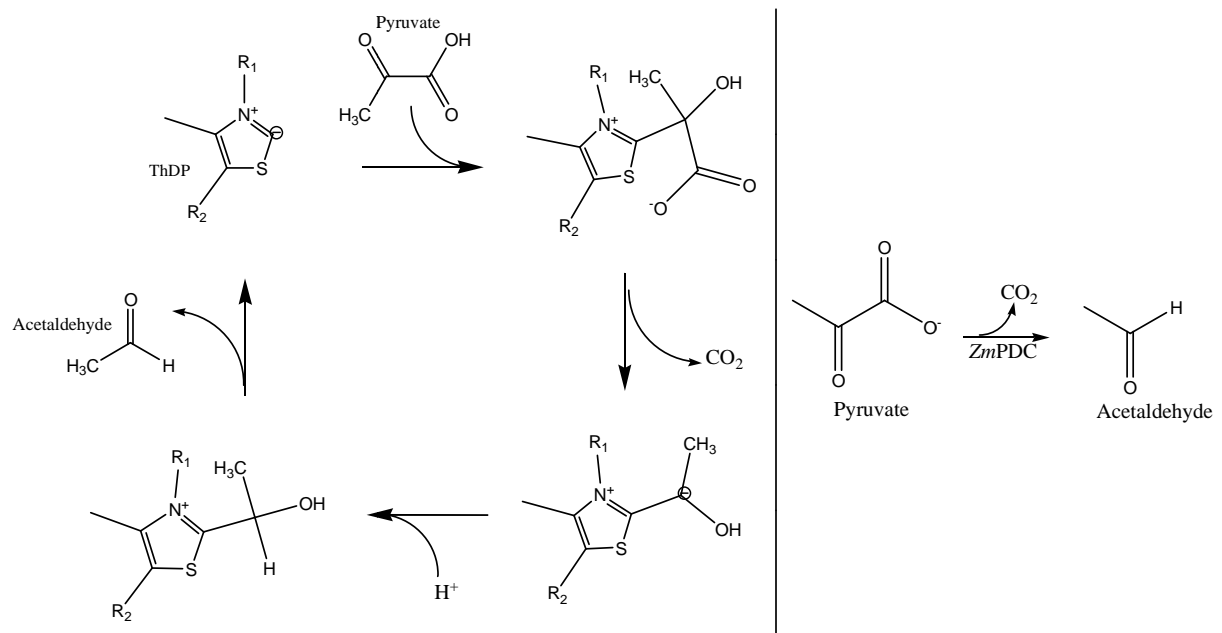


Figure 1.15 – The catalytic cycle of *ZmPDC* and the general reaction that produces acetaldehyde from pyruvate via *ZmPDC* catalysis are indicated.

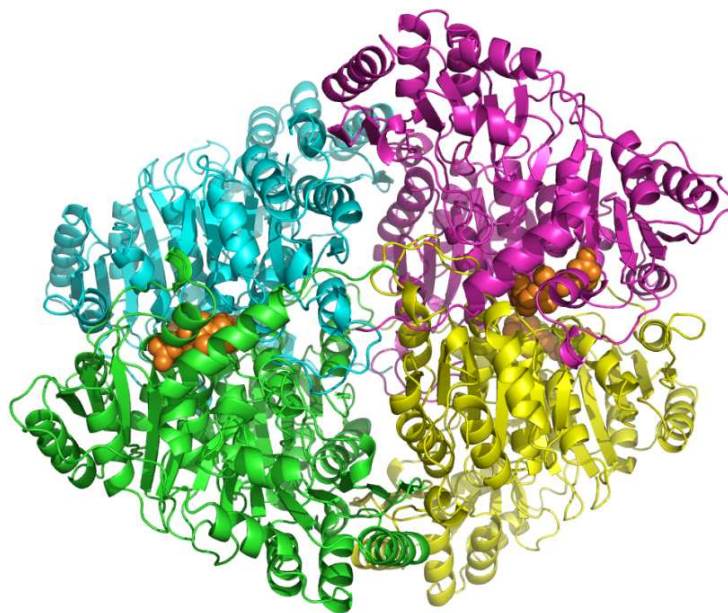


Figure 1.16 - *ZmPDC* model (PDB code: 1ZPD) with ThDP shown as orange spheres

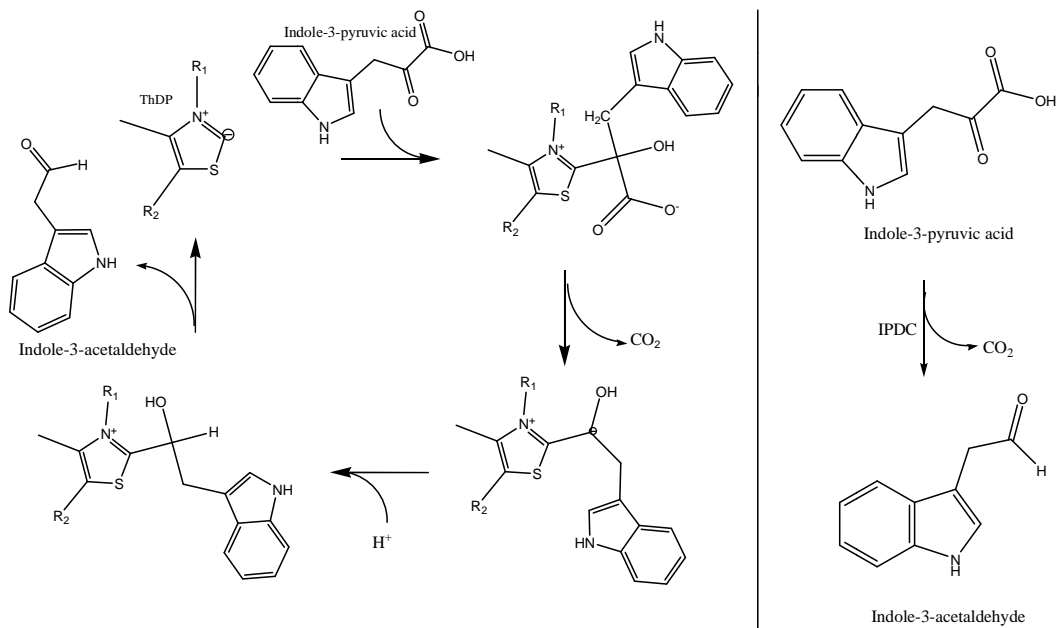


Figure 1.17 - The catalytic cycle of IPDC and the general reaction that produces indole-3-acetaldehyde from indole-3-pyruvic acid via BFD catalysis are indicated.

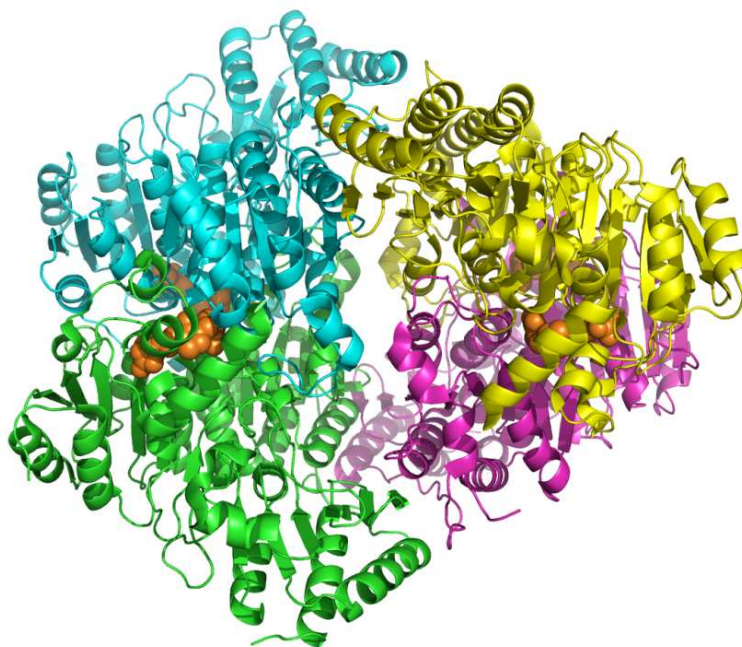


Figure 1.18 - IPDC model (PDB Code: 1OVM) with ThDP depicted as orange spheres

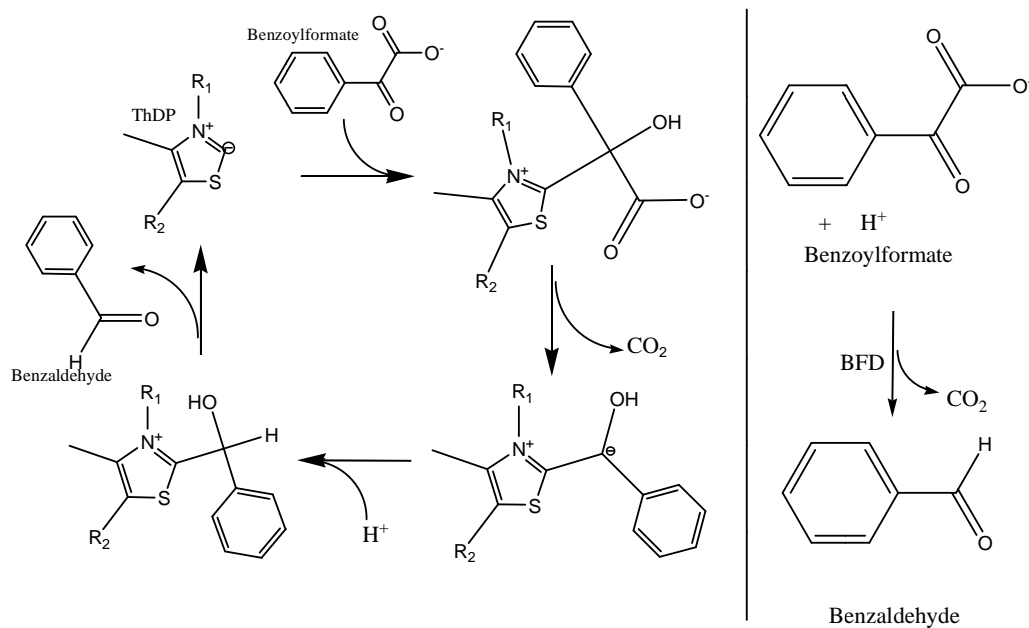


Figure 1.19 - The catalytic cycle of BFD and the general reaction that produces benzaldehyde from benzoylformate via BFD catalysis are indicated.

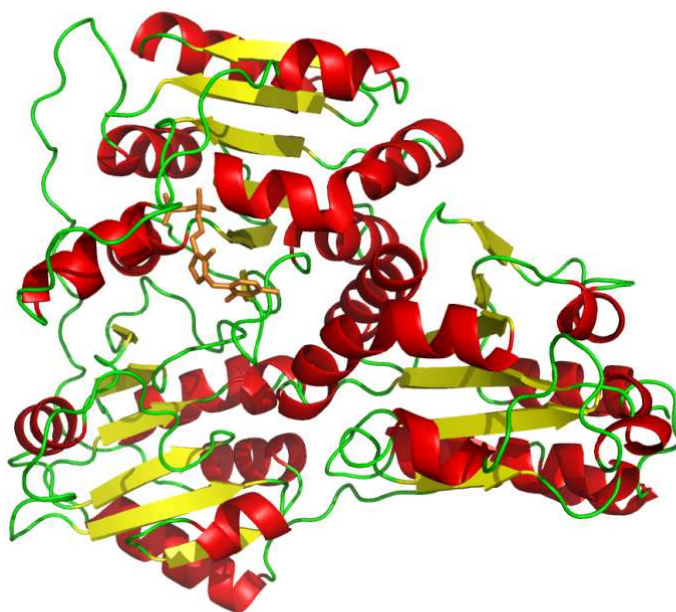


Figure 1.20 – BFD Model (PDB Code: 1BFD). ThDP is depicted in orange as a stick model. The monomer is shown as presented within the protein data bank

1.5.2 DC Family enzymes that utilize the same or similar primary and secondary substrates

Another subset of enzymes within the DC family exists where a secondary substrate similar to the primary substrate is attacked by the enamine/carbanion intermediate. Three enzymes with structures elucidated fall into this category: acetohydroxy acid synthase (AHAS) (Pang 2002), acetolactate synthase (ALS) (Pang 2004), and benzaldehyde lyase (BAL) (Mosbacher 2005).

AHAS catalyzes the formation of acetohydroxy acid from pyruvate and 2-ketobutyrate (Figure 1.21). ALS forms acetolactate from two molecules of pyruvate (Figure 1.23), a reaction that AHAS can complete as well (Figure 1.21). Despite performing similar reactions the structures of the two enzymes are not identical. Structural differences within the active site make the binding of 2-ketobutyrate, which is seen in AHAS, sterically impossible in ALS. Furthermore, AHAS requires the binding of FAD as a cofactor whereas ALS does not utilize FAD or have a nucleotide binding site. Catalytically the enzymes are similar with both having a methionine, glutamine and lysine residue participating in catalysis (AHAS: Lys251, Gln 202 and Met582; ALS: Lys36, Gln420, Met 479) (Pang 2004). Sequence alignment of AHAS and ALS shows that the glutamine and lysine residues do not align and are in fact from different domains of the protein that form the active site. Although ALS and AHAS are thought to have a common ancestor (Bowen 1997) the two enzymes seem to have evolutionally converged to share similar functionality.

Benzaldehyde lyase (BAL) catalyzes the stereospecific formation of *R*-benzoin from two benzaldehyde molecules, as well as the irreversible cleavage of *R*-benzoin (Figure 1.25). BAL

also accepts a variety of other substituted aromatic or aliphatic acyl-acceptors as substrates and has therefore been thought to be a potential tool for chemo-enzymatic synthesis of enantiomerically pure 2-hydroxy ketones (Demir 2001). For BAL both the forward and the reverse reactions are thought to utilize a water molecule which is hydrogen bonded to His29 and Gln119 to perform the catalytic reaction with the substrate.

Sequence alignment and structural analysis indicate that BAL maintains the GDG-(X₂₄)-NN motif that is conserved for binding of ThDP and Mg²⁺. However, this is not the case for ALS and AHAS. ALS does not contain the NN portion of the GDG-(X₂₄)-NN motif, but rather has a valine and aspartate residue in their stead. A mutation of the second asparagine in other ThDP enzymes has proven detrimental to ThDP binding. For example, in pyruvate decarboxylase from *Z. mobilis* the mutation of the asparagine residue (Asn467Asp) resulted in drastic loss of affinity for ThDP with the concentration of ThDP required for half saturation of the enzyme (S_{0.5}) increasing by nearly three orders of magnitude (Candy 1998). For AHAS the binding motif is altered to GDA-(X₂₄)-NN. This alteration is insignificant since the residue side chain is not involved in binding of ThDP or Mg²⁺ at this location but rather the main chain oxygen.

The first structure within this subgroup of DC family enzymes to be solved was that of AHAS (Figure 1.22) (Pang 2002). The structure was solved via molecular replacement utilizing BFD as the search model. The same research group then published the structure of ALS (Figure 1.24) two years later by using the AHAS catalytic residues for molecular replacement (Pang 2004). Even with several DC family enzyme structures published, the structure of BAL was not solved with molecular replacement. Rather, a SeMet derivative of BAL was used to solve the structure (Figure 1.26) via multiple wavelength anomalous dispersion (MAD) (Mosbacher 2005).

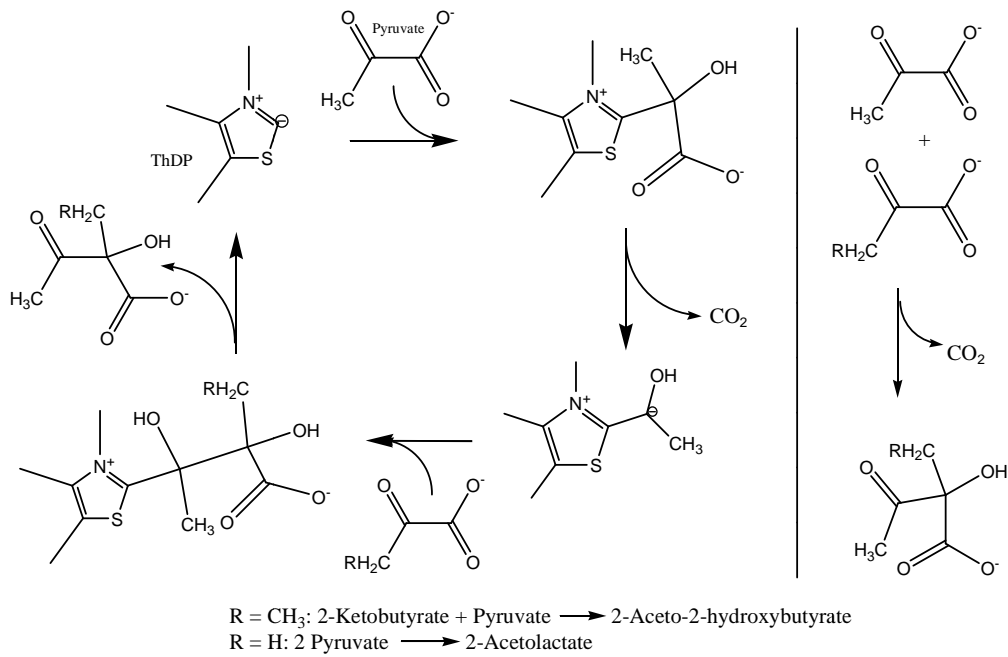


Figure 1.21 - The catalytic cycle of AHAS. This enzyme can catalyze two separate reactions to produce 2-acetolactate and 2-aceto-2-hydroxybutyrate

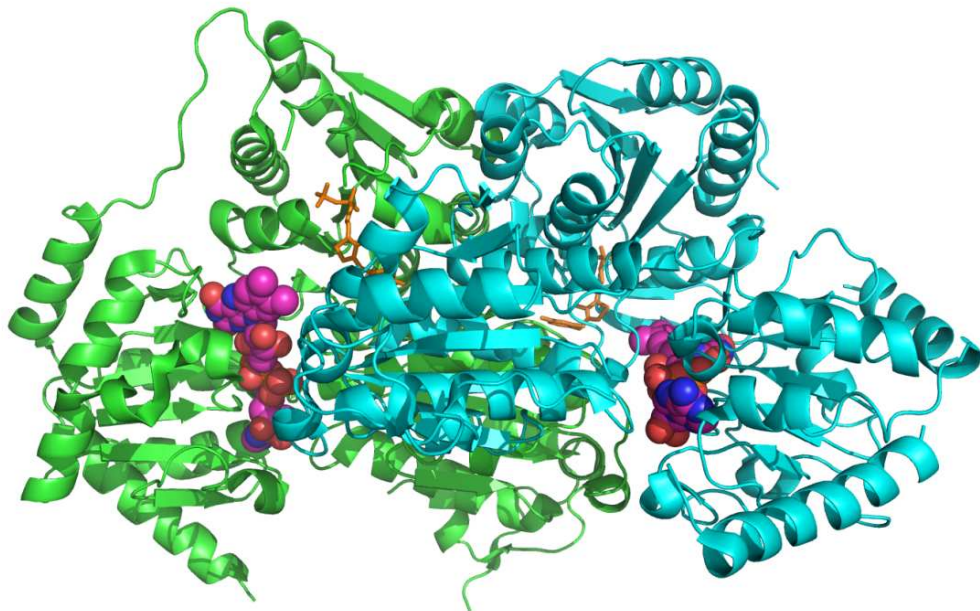


Figure 1.22 - AHAS model (PDB Code: 1JSC). FAD is depicted as spheres and ThDP is depicted in orange as sticks.

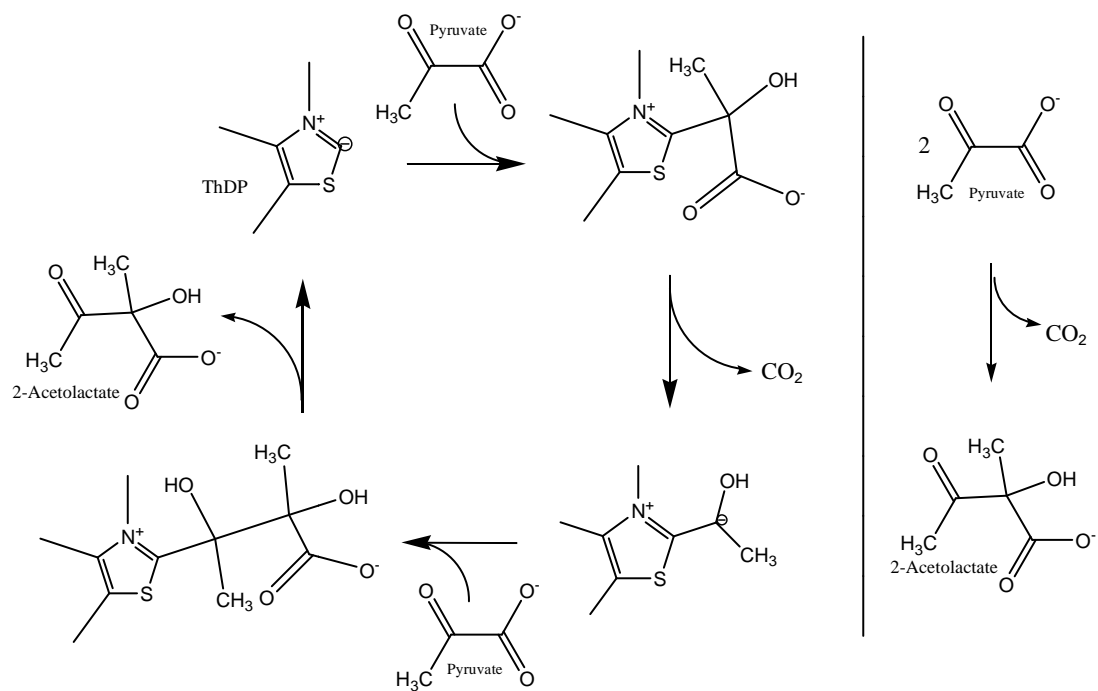


Figure 1.23 - The catalytic cycle of ALS and the general reaction that indicates the production of 2-acetolactate from two molecules of pyruvate

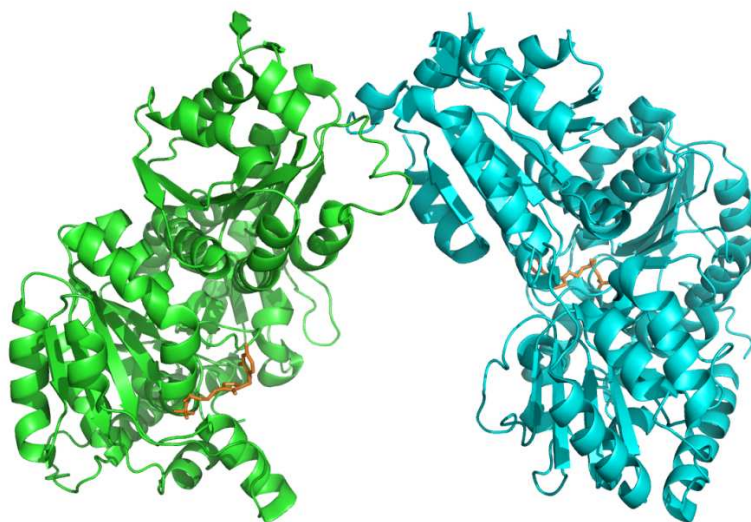


Figure 1.24 - ALS model (PDB Code: 1OZG) with ThDP depicted in orange as a stick model; note the absence of FAD in the structural representation.

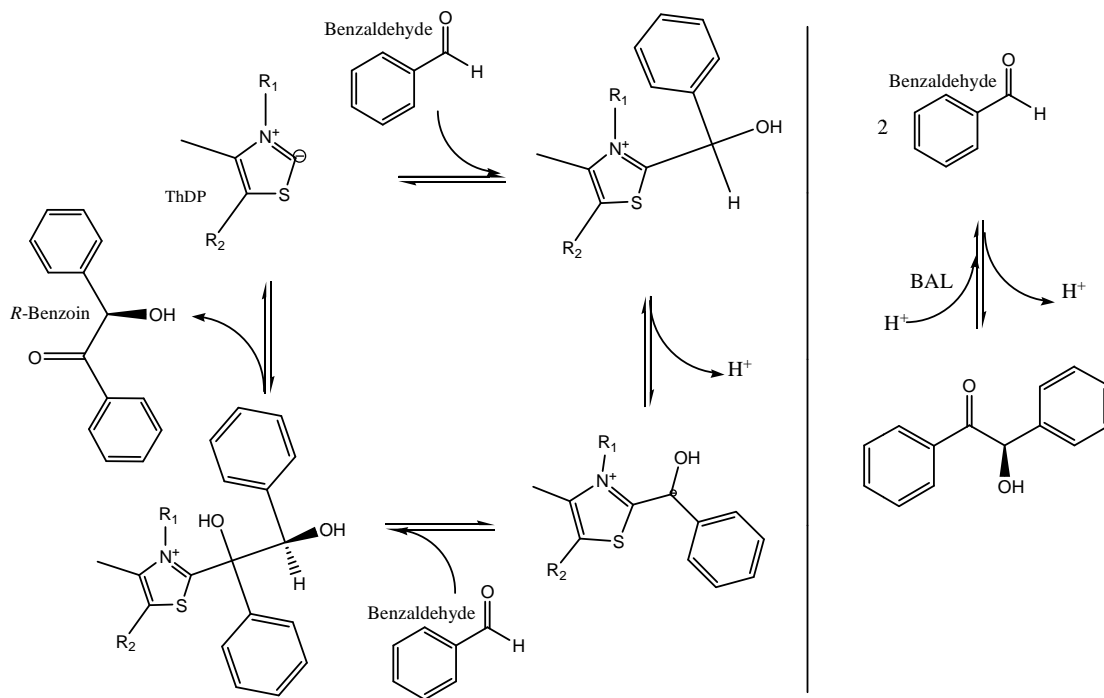


Figure 1.25 - BAL catalytic cycle indicating the reversible cleavage and formation of *R*-benzoin to/from two molecules of benzaldehyde

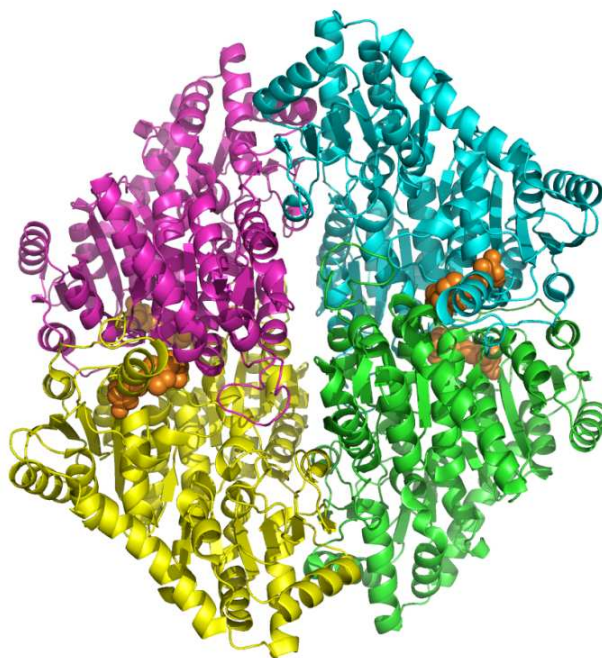


Figure 1.26 – BAL model (PDB Code: 2AG0). ThDP is indicated with orange spheres.

1.5.3 Other DC Family Enzymes with Elucidated Structures

1.5.3.1 Carboxyethyl Arginine Synthase

Carboxyethyl arginine synthase catalyzes the dephosphorylation-carboligation reaction of D-glyceraldehyde and L-arginine to form N^2 -(2-carboxyethyl)arginine (Figure 1.28) (Khaleeli 1999) The reaction is believed to follow the general transferase reaction mechanism with the aldehyde functional group of D-glyceraldehyde being activated, allowing for nucleophilic attack on L-arginine.

The structure of CEAS was determined using multiple wavelength anomalous dispersion (MAD) using a SeMet mutant protein (Caines 2004). Based on the crystal structure with a sulphate ion bound to the active site in place of D-glyceraldehyde it is believed that this substrate is bound by two histidine residues (H120, H415) (Caines 2004). These two histidines do not participate in hydrogen donation or abstraction like is seen commonly with histidine pairs in the active site or ThDP dependent enzymes. It was assumed that much of the general acid-base catalysis is completed by ThDP itself (Khaleeli 1999). The structure of the enzyme indicates that the active site residues are not capable of general acid-base catalysis and that the 4'-amino group of ThDP plays an important role in catalysis (Caines 2004).

The primary sequence and structure of CEAS indicates that the GDG-(X₂₄)-NN ThDP binding motif and the glutamate required for activation of ThDP are in place. The sulphate ion bound to CEAS is thought to be comparable to the phosphate of D-glyceraldehyde. Also, this sulphate is within proper distance from ThDP to allow for the four carbon bond links of D-glyceraldehyde as well (Caines 2004).

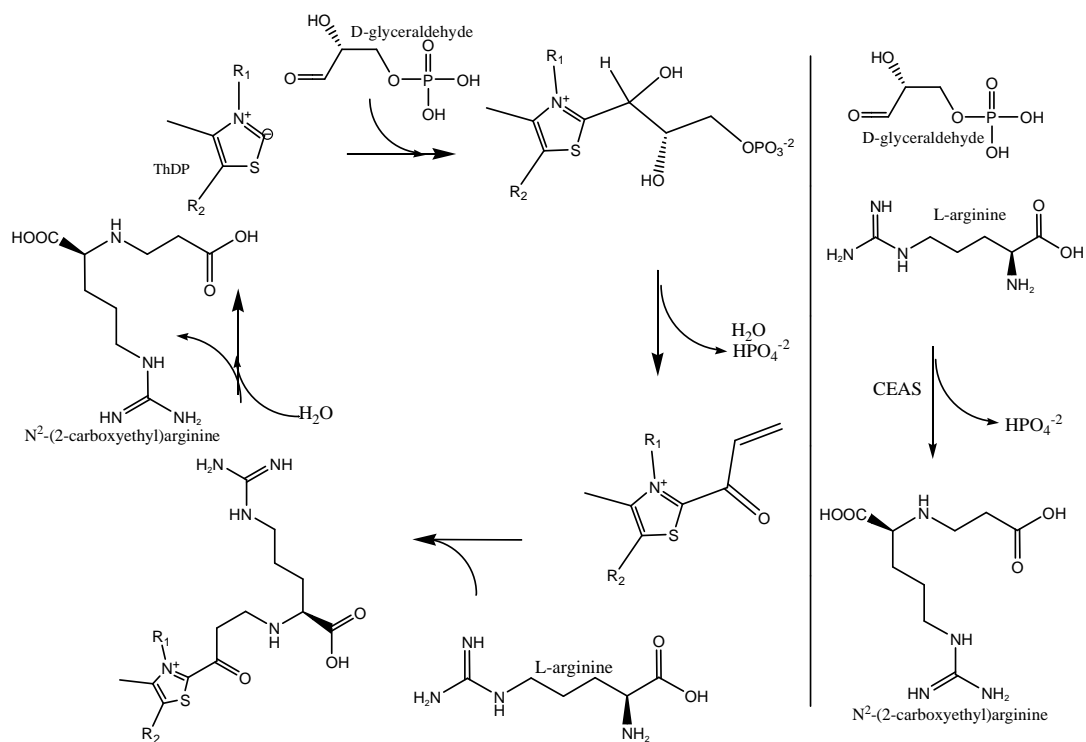


Figure 1.27 - CEAS catalytic cycle indicating the formation of N²-(2-carboxyethyl)arginine from D-glyceraldehyde and L-arginine

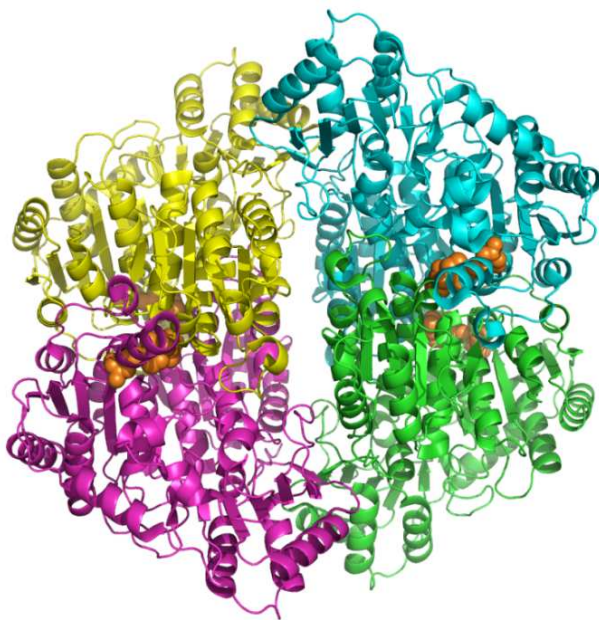


Figure 1.28 - CEAS model (PDB Code: 1UPA). ThDP in indicated with orange spheres

1.5.3.2 Pyruvate Oxidase

Pyruvate oxidase catalyzes both an oxidative decarboxylation and a phosphorylation reaction, in which, pyruvate is converted to acetylphosphate (Figure 1.26) (Gotz 1991). Unlike any of the other enzymes in the DC Family, POX utilizes a second cofactor, FAD, for oxidative alterations to the primary substrate (Figure 1.29). The enzyme still follows the general decarboxylation mechanism for ThDP dependent enzymes, yet it is the only ThDP dependent enzyme to utilize a phosphate molecule as a secondary substrate and also have a tertiary substrate, oxygen. The transfer of two electrons from oxyethyl-ThDP to FAD is facilitated by the presence of two phenylalanine residues from different domains (Phe121 and Phe479). FAD is then regenerated by oxygen to produce hydrogen peroxide. Subsequently, an inorganic phosphate group attacks the intermediate, releasing acetylphosphate and regenerating ThDP (Muller 1994).

The structure of POX (Figure 1.30) was determined using multiple isomorphous replacement (MIR) (Muller 1994). Phasing was initialized by utilizing data from seven heavy atom derivatives and an initial electron density map refined to 3.5 Å was produced. The structure of the native enzyme was then determined to 2.1 Å resolution utilizing difference Fourier maps, iterative model building and refinement.

The conserved binding motif is altered to GDG-(X₂₄)-TN within POX. Similar to ALS the primary interaction by the asparagine portion of the motif is altered. However, as is indicated in Figure 1.12 the first asparagine residue side chain is not directly involved in ThDP or Mg²⁺ binding. Therefore, this alteration is again not very substantial and as expected threonine is not implicated in ThDP binding (Muller 1994).

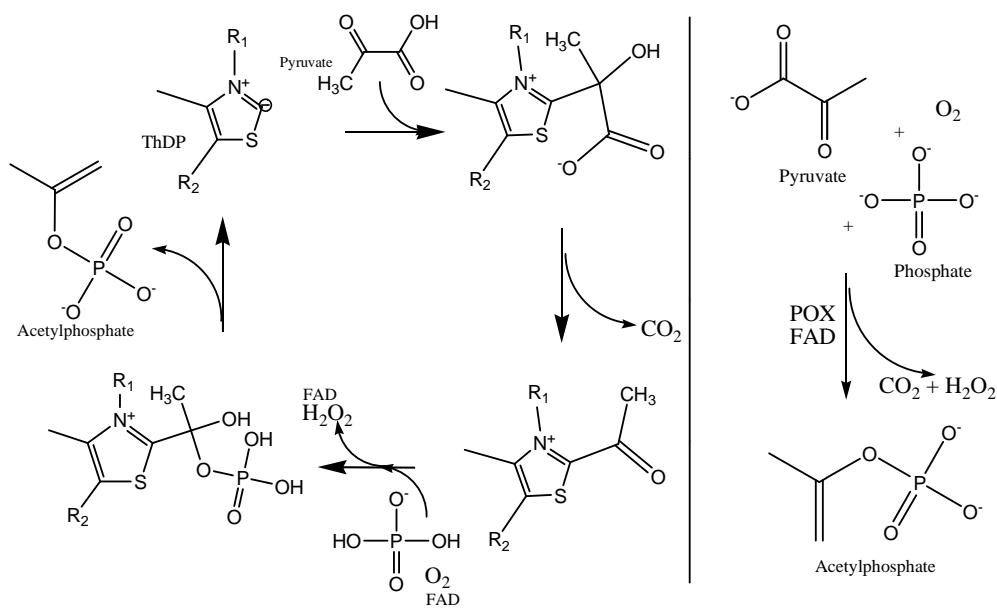


Figure 1.29 - POX catalytic cycle indicating the formation of acetylphosphate from pyruvate, oxygen and inorganic phosphate; along with the reduction of FAD

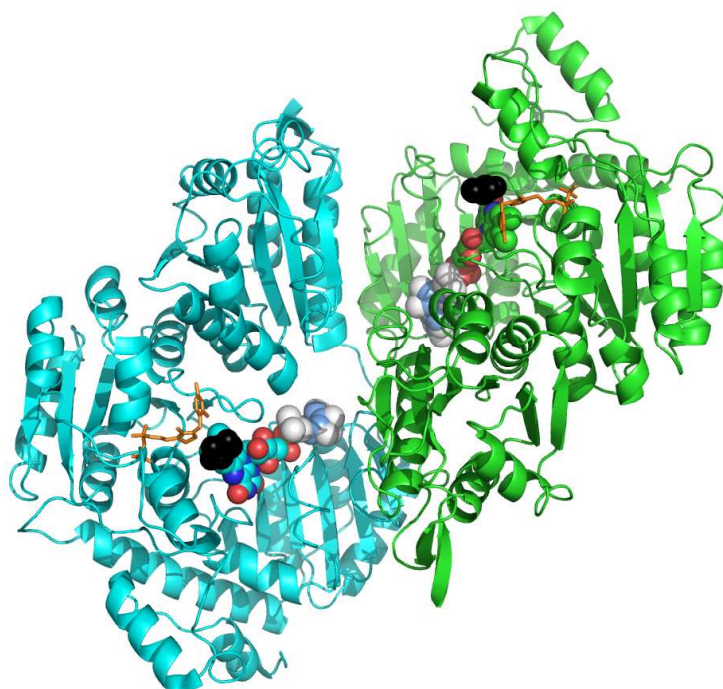


Figure 1.30 - POX Model (PDB Code: 1POX). FAD is depicted as spheres and ThDP is indicated with orange stick models

1.5.3.3 Oxalyl-CoA Decarboxylase

Oxalyl-CoA decarboxylase (OCDC) catalyzes the non-oxidative decarboxylation of oxalyl-CoA to formyl-CoA (Figure 1.31). OCDC follows the general decarboxylation mechanism with H^+ being the secondary substrate for the completion of the reaction. This is a similar mechanism seen with the PDC's, IPDC and BFD.

The structure of OCDC (Figure 1.32) was solved using molecular replacement techniques utilizing a polyalanine model of AHAS as the successful search model (Berthold 2005). Unlike the other enzymes which utilize H^+ as a secondary substrate, OCDC does not have a pair of histidine residues present within the active site for catalysis. Rather OCDC contains a hydrogen bond network between a water molecule, the main chain nitrogen of Ile34, and the side chains of Tyr120 and Glu121. These residues contribution to catalysis were confirmed with co-crystal structures containing substrate (Berthold 2007).

There are some differences structurally for this enzyme when compared to other ThDP-dependent enzymes. As is quite evident from the reaction scheme in Figure 1.31, the substrate for this enzyme is considerably larger than most. The much larger substrate is accommodated by a binding pocket that bears resemblance to that seen in ThDP-dependent enzymes that contain FAD as a cofactor. Further differences are seen as OCDC also shows an allosteric effect from the addition of adenine diphosphate (ADP) (Berthold 2005). Allostery is not noted within any of the other ThDP dependent enzymes. Also, the conserved ThDP binding motif is altered to GDS-(X₂₄)-NN within OCDC. Similar to AHAS this alteration is thought to be insignificant since the main chain carbonyl contributes to ThDP and Mg^{2+} binding and not the side chain.

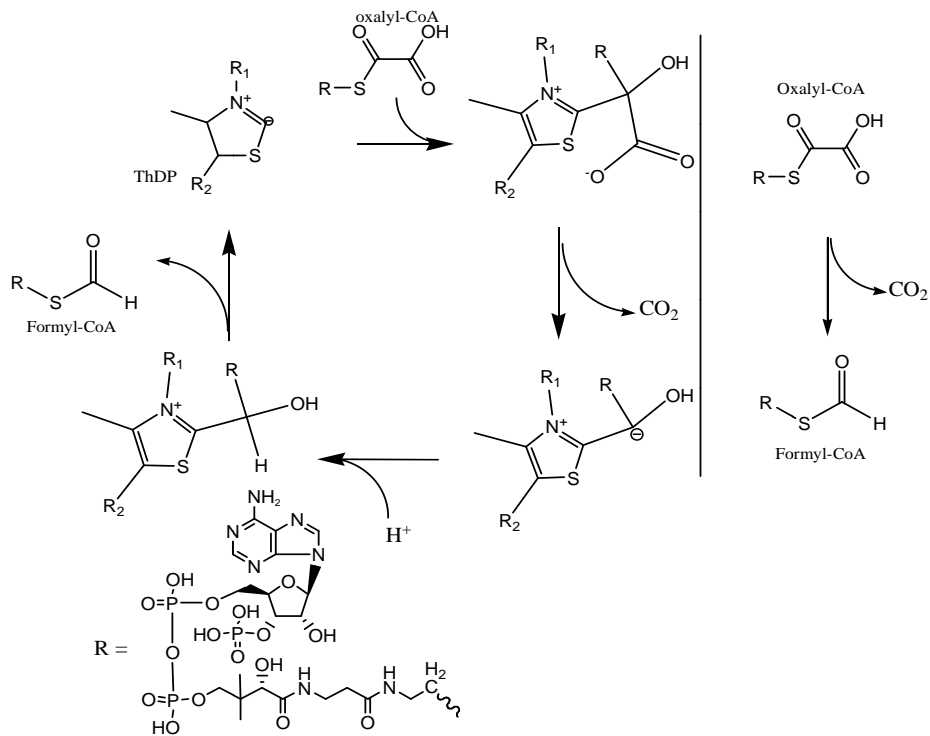


Figure 1.31 - OCDC catalytic cycle indicating the formation of formyl-CoA from oxalyl-CoA.

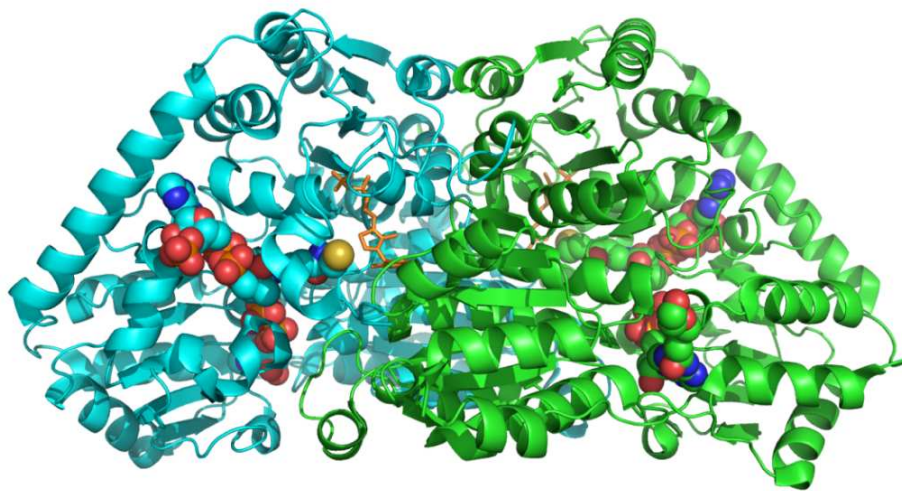


Figure 1.32 - OCDC Model (PDB Code: 2JIB). ADP and CoA are represented as spheres and ThDP depicted as an orange stick models

1.6 MenD

As stated within section 1.4, MenD is also believed to be part of the DC family of enzymes (Duggleby 2006). However, the structure of *E. coli* MenD has not been elucidated and no theoretical model has been generated for this enzyme. This leaves a significant gap in the understanding of MenD's structure-function relationship, catalytic processes and information necessary to develop inhibitors.

Previous work done within the Palmer lab has implicated a variety of residues as required for catalysis. Mutagenesis and subsequent kinetic analysis has allowed for a variety of residues to be implicated as contributors to MenD catalysis, ThDP binding and activation, and also Mg^{2+} binding. Among them is the conserved glutamate residue, Glu55 in MenD, which is thought to be responsible for ThDP activation, since even conservative mutagenesis such as Glu55Asp results in complete loss of enzyme activity (Macova 2005).

Another residue implicated is Ile418, which is thought to aid in the induction of the V-conformation of ThDP upon binding. Again, a conservative mutation such as Ile418Leu results in near complete loss of enzymatic activity. One of the residues thought to be required for catalysis is Ser32, since mutagenesis has shown that the variant Ser32Asp results in complete loss of activity (Ho 2007).

A variety of other residues have been investigated with mutagenesis studies, including Asp442, which was implicated for Mg^{2+} binding within the ThDP-dependent enzyme binding motif. As well, Arg33, Arg395, Gln118 are all thought to participate in catalysis (Macova 2005).

The use of crystallographic methods will allow the determination of the actual interaction of all of these residues and their role in MenD catalysis. The information provided from structural analysis will provide a far more complete understanding of MenD, its associated mechanism, and its relation to various homologous enzymes. Furthermore, the elucidation of a protein structure will allow for the development of inhibitors for this enzyme that could potentially serve pharmaceutical purposes.

Chapter 2: Research Goals and Approach

My primary research goal was to solve the phase problem from a MenD X-ray diffraction data set. Once the phase problem was solved, a structural model of MenD could be determined, refined and submitted to the protein data bank.

In August of 2005, Edyta Sieminska grew crystals of MenD protein expressed from pMD14b. Two diffraction data sets were collected by Edyta at the Advanced Photon Source in Chicago later that month from two different crystal types. The first experiment produced a high resolution (1.9 Å) diffraction data set from native holoenzyme MenD crystals (co-crystallized with Mn^{2+} and ThDP). The second was a medium resolution (2.4 Å) diffraction data set collected from apoenzyme crystals.

My work began with the analysis and processing of these data sets and the goals of this work were as follows:

- 1) Process both diffraction data sets with MOSFLM and SCALA
- 2) Determine which of the data sets is more suitable for molecular replacement strategies
- 3) Solve the phase problem with molecular replacement software programs
- 4) Produce/refine a structural model of MenD to aid in the understanding of the enzyme structure-function relationship.

If these methods were unsuccessful, my research was to focus on the production of diffractable MenD protein crystals that contain seleno-methionine (SeMet) residues rather than methionine residues. This was thought to allow for the elucidation of the structure of MenD

using multiple wavelength anomalous dispersion (MAD) techniques with selenium acting as a heavy atom derivative. The steps that were outlined to achieve this goal are listed below:

- 1) Clone the *E. coli* MenD gene into an appropriate vector for SeMet expression
- 2) Show that the new construct expresses active MenD protein
- 3) Provide experimental evidence that the new construct can be expressed in minimal media containing methionine
- 4) Show that the MenD from the new construct can be purified using buffers containing a reducing agent such as TCEP
- 5) Produce diffractable native crystals of MenD expressed from the new construct
- 6) Produce SeMet derivative protein by expressing in minimal media devoid of methionine and containing SeMet
- 7) Show that SeMet protein can be purified using the same procedure employed with native protein in the presence of TCEP
- 8) Test level of inclusion of SeMet into protein using mass spectroscopy
- 9) Produce diffraction quality crystals of SeMet derivative
- 10) Collect full data set at three separate wavelengths at the Advanced Photon Source in Chicago for the SeMet derivative of MenD
- 11) Solve the phase problem utilizing MAD computational methods to identify the location of SeMet residues and refine the structure

Chapter 3: Materials and Methods

3.1 General

Luria-Bertani (LB) broth used for growth of bacterial cultures followed the Miller formulation, utilizing ten grams of Bacto™ tryptone (Becton, Dickinson and Company), ten grams of sodium chloride (EMD Chemicals Inc.) and five grams of Bacto™ yeast extract (Becton, Dickinson and Company) per liter of water. Likewise, LB-agar followed the same formulation along with the addition of fifteen grams of granulated agar-agar (EM Science). The prepared solutions were then autoclaved prior to addition of ampicillin. Upon cooling to 50 °C the LB-agar solution had ampicillin added to it and then was poured into sterile polystyrene petri dishes (VWR) in twenty milliliter portions.

3.2 Software Utilized

Molecular replacement was attempted with various software packages including AMoRe (Navaza 1994), Phaser (Read 2001) and MR BUMP (Keegan 2006). The data which was used for molecular replacement was integrated and scaled using MOSFLM (Leslie 1992) and SCALA (Weiss 2001). Polyalanine models were produced with the use of the MOLEMAN2 (Kleywegt 1998). A MenD model was generated with the program Modeller (Fiser 2000) using 2AG0 as a reference model. Least squares analysis of the MenD model to 2AG0 was completed using LSQMAN (Kleywegt 1996). Structure based sequence alignments were performed using the online program FUGUE (Shi 2001) and the output is reformatted using ESPript (Gouet 1999).

PDB files are analyzed with Coot (Emsley 2004) and presented within this document using PyMol (DeLano 2002).

3.3 Cloning Procedures

Primers were designed that could allow for the copying of the MenD gene from the pMD14b plasmid (developed from pET14b) via the polymerase chain reaction (PCR) and also included restriction sites that were appropriate for fusion into pQE80L. These primers were a forward BamHI primer with the sequence 5' CCGCGCGGATCCCATATGC 3' and the reverse KpnI primer with sequence 5' GCAGCCGGTACCTCATAAATGGC 3' (provided by Alpha DNA). A PxE 0.2 Thermocycler (Thermo Electron Corporation) was then used to perform a PCR. The PCR program utilized a 50 μ L mixture containing 93 ng / μ L pMD14b plasmid (93.4 ng / μ L, 2.0 μ L), 20 μ M forward BamHI primer (1.0 μ L), 20 μ M reverse KpnI primer (1.0 μ L), 25 μ M dNTP (0.4 μ L), nuclease free double distilled water (39.6 μ L), Taq polymerase (1.0 μ L), and 10x PCR buffer (5.0 μ L). An annealing temperature of 56 $^{\circ}$ C was decided upon given the provided melting point temperatures from Alpha DNA. The solution was first heated to 95 $^{\circ}$ C for two minutes to ensure melting of all plasmid DNA. The temperature was then dropped to 52 $^{\circ}$ C to ensure strong annealing of the primers to the strand and then the bound primers were extended by adjusting to the optimal *Taq* polymerase temperature of 72 $^{\circ}$ C for two minutes. The following cycle was then repeated thirty times: melting at 95 $^{\circ}$ C for thirty seconds, annealing at 56 $^{\circ}$ C for one minutes and then extension at 72 $^{\circ}$ C for the next two minutes (total time of 30 cycles was 105 minutes). The solution was then placed at 72 $^{\circ}$ C for an additional ten minutes and then the temperature reduced to 4 $^{\circ}$ C until the system was shut down.

The product of this reaction was then run on a 1% agarose gel and a band of the appropriate predicted size was identified on the gel with comparison to 1 Kb Plus DNA ladder (Invitrogen). This band was then excised and purified using a QIAquick gel extraction kit (Qiagen) using the outlined procedure in the provided manual. The purified DNA insert was then doubly cut with KpnI and BamHI using a solution containing 2.0 μ L KpnI, 2.0 μ L BamHI, 12.0 μ L of purified DNA, 2.0 μ L of 10x Buffer and 2.0 μ L of H₂O. Similarly, pQE80L plasmid was cut with KpnI and BamHI. Nova Blue cells were then transformed with a ligation mixture using the heat shock method with a forty-five second exposure to a 42 °C bath with immediate return to an ice bath. The transformed cells were then plated on LB Amp⁺ plates with the most colonies occurring in solution prepared with a 5:1 ratio of cut insert DNA concentration to cut pQE80L DNA for the ligation.

3.4 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

In the case of MenD, a 61.4 kilodalton (kDa) protein as a monomer, a 10% acrylamide gel was prepared. A 10% acrylamide gel allows for separation of proteins between twenty and eighty kDa. The gel was prepared with two components, a stacking gel solution and a separating gel solution, in accordance with a standard recipe (Table 2.1).

Table 3.1 - SDS-PAGE gel preparation recipe for one gel (*APS must be made fresh)

Component	10% Separating Gel	Stacking Gel
H ₂ O	3.55 mL	1.8 mL
1.5M Tris (pH 8.8)	1.9 mL	0.315 mL
40% Acrylamide mix	1.9 mL	0.315 mL
10% SDS	75 μ L	25 μ L
10% APS*	75 μ L	25 μ L
TEMED	7.5 μ L	2.5 μ L

Materials used to prepare these two solutions were all purchased from Sigma-Aldrich with the exception of ammonium persulphate, which was purchased from VWR International Ltd. The dual gel casting unit and associated parts was purchased from Hoefer Scientific.

In order to identify MenD on the acrylamide gel a molecular weight standard was also loaded (SigmaMarker™ Low Range (M.W. 6,500-66,000)). The electrophoresis experiment was run at a constant voltage (220 V) until the tracking dye reached the bottom the gel. The gel was then removed from the casting unit and stained in coomassie brilliant blue (VWR Scientific). After staining is complete the gel is destained using two solutions, a fast destaining solution (30% methanol, 10% acetic acid), and a slow destaining solution (5% methanol, 7% acetic acid). Finally, the gel is dried on a gel drying film (VWR Scientific) using drying solution (40% methanol, 10% glycerol, 7.5% acetic acid).

3.5 Expression Procedures

3.5.1 Protein Expressed from pMD80L plasmid

Utilizing a single pMD80L transformed colony from a LB^{Amp}-agar plate, a 10 mL LB broth culture was grown overnight at 37 °C in the presence of 100 µg/mL ampicillin. The next morning 1.0 mL of this solution was added to 1.0 L of autoclaved LB broth containing 1.0 mL of 100 mg/mL ampicillin. The solution was then placed in an Orbit environ shaker at 225 rpm and the growth of cells was monitored with the progression of optical density at 600 nm (OD₆₀₀) using a UV/Vis spectrometer. Upon obtaining a OD₆₀₀ that falls in the range of 0.4 to 0.8, the temperature of the shaker was lowered to 30 °C and the cells were induced with 1.0 mL of 0.1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). Following four hours of growth the cells

were harvested via centrifugation at 8000 rpm in a Beckman J2-HS centrifuge using a JLA 10,500 rotor.

The pelleted cells were collected and then resuspended in cold lysis solution. A 25x lysis solution was composed of 25 mM DNase, 25 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 1.4 mM lysozyme. 1.0 mL of this 25x lysis solution was added to 12.0 mL of cells suspended in 12.0 mL of 50 mM Tris-HCl pH 7.8-7.9. The final concentration of Tris-HCl was thusly 25 mM and the lysis buffer was diluted to 1x. This mixture was kept cold in an ice bath or placed in a 4 °C fridge and mixed vigorously with a stir bar for thirty minutes.

Following completion of exposure to the lysis solution the cell solution was then sonicated with a Virtus Virsonic ultrasonic cell disruptor. Volumes greater than 35 mL used the larger probe, whereas smaller volumes used the micro-tip probe. The total time of sonication was three minutes with intervals set up to have the sonicator activated for two seconds and then deactivated for three seconds. The solution was kept on ice. The three minute total time was split into three one minute intervals with two minutes between each interval to allow the solution to cool.

3.5.2 Seleno-methionine Protein Expressed from pMD80L

Expression of selenomethionine (SeMet) containing protein was accomplished with the utilization of an expression kit (MD045004-12L) provided by Medicilon. This kit utilizes methionine pathway inhibition to ensure high inclusion of SeMet and minimal amounts of methionine within the protein. The kit was used for the growth of a 1.0 L culture of SeMet protein. However, the procedure utilized for the inoculation stage did not follow the outlined

protocol. Instead, 10.0 mL of the cells were grown in the standard conditions for native production of pMD80L MenD for inoculation. The cells were then centrifuged at 15000 rpm in a Beckman J2-HS centrifuge containing a JA 25.50 rotor. The supernatant was then poured off and the pellet re-suspended in 10.0 mL of 1x M9 salts solution. The M9 solution used was prepared as a 10x solution containing 60.0 grams of Na₂HPO₄, 30.0 grams of KH₂PO₄, 10.0 grams of NH₄Cl, 5.0 grams of NaCl and then double distilled water was added to a volume of one liter. This mixture was then autoclaved. The resuspended cells were then added to the expression kit as directed in a 10.0 mL aliquot.

The cells were grown to an OD₆₀₀ of 1.2 and then the temperature of the solution was decreased to 20 °C prior to the addition of the inhibitor cocktail and the SeMet. The solution was then allowed to grow overnight (approximately twelve hour incubation) and was then harvested.

3.6 Purification Procedures

3.6.1 Purification of *E. coli* MenD From pMD80L Construct

The procedure utilized for the purification of the protein produced from the expression of pMD80L construct involved the utilization of a POROS MC-50 His-tag column. The purification was performed using a BIOCAD 700E perfusion chromatography workstation and an Advantec SF-2120 superfraction collector.

The first stage involved equilibration of the column using five column volumes of filtered water (40 mL / min), followed by ten column volume of 50 mM NiSO₄ (10 mL / min) to charge the column, and subsequent washing with another five column volumes of water (40 mL / min).

Prior to loading the column with protein sample the column was exposed to five column volumes of binding buffer. A 1.0 L solution of the binding buffer contained 5 mM imidazole, 0.5 M sodium chloride, 12.5% glycerol, and 20 mM Tris-HCl, pH 7.9. The final pH of the solution was then adjusted to pH 7.9 with the addition of 1 M NaOH.

Following equilibration of the column with binding buffer, 5.0 mL of the sample was injected onto the column by the BIOCAD and then the unbound protein was washed away with another five column volumes of binding buffer. The system then repeats the injection of another 5.0 mL of protein sample and another washing with five column volumes of binding buffer.

Following the second injection the column is washed with ten column volumes (40 mL / minute) of wash buffer to elute non-specifically bound protein from the column. A 1.0 L solution of wash buffer contains 50 mM imidazole, 0.5 M sodium chloride, 20 mM Tris-HCl pH 7.9, and 12.5% glycerol. The final pH of the solution was again adjusted to 7.9 with 1 M NaOH.

Upon completion of the wash stage the bound protein is eluted with a gradient from 100% wash buffer to 100% elution buffer over fifteen column volumes at (40 mL / min). While the elution of the protein was occurring the Advantec SF-2120 superfraction collector was collecting 10 mL fractions (12 in total). A 1.0 L solution of the elution buffer contains 0.1M ethylenediamine tetraacetic acid (EDTA), along with 0.5 M sodium chloride, 20 mM Tris-HCl pH 7.9, and 12.5% glycerol. The final pH was once again adjusted to pH 7.9 with a large amount of NaOH in order for the EDTA to dissolve in solution (See Figure 4.14 for sample purification output).

The column is cleaned for next usage with an additional five column volumes of elution buffer, as well as ten column volumes of each of 1.0 M HCl, 0.5 M NaOH and finally double distilled water. The entire column cleaning procedure is completed with a flow rate of 40 mL per minute.

3.6.2 Purification of SeMet *E. coli* MenD

The purification procedure for the seleno-methionine derivative was the same as that for the wild-type pMD80L protein with a single alteration. Each of the binding, wash and elution buffers contained Tris(2-carboxyethyl) phosphine (TCEP) at a concentration of 2 mM. The addition of this compound was deemed necessary for the SeMet construct since the presence of a strong reducing agent ensure that the selenium within the protein does not become oxidized.

3.7 Dialysis Procedures

Dialysis procedures were completed with the purified protein to remove residual components from the purification procedure. A three step dialysis procedure was carried out by placing the purified protein solution within Spectra/Por Dialysis Membrane (VWR Scientific) with a molecular weight cut off of 6-8 kDa. The dialysis solution was one liter in volume and varied slightly depending on the components included to have either holo or apo enzyme. In the case of the apoenzyme the solution consisted of 2 mM MgCl₂, 150 mM NaCl, 50 mM Tris-HCl at pH 8.0 and 10% glycerol. For the holoenzyme the solution remained the same, however ThDP was also added to the mixture at a concentration of 100 μM. In the case of SeMet protein the addition of 2 mM TCEP was also required to ensure that the selenium was not being oxidized, as was the case with purification as well.

In the latter stages of this project it was discovered that the 10% glycerol was not necessary to be added to prevent protein precipitation. As such, future dialysis of the protein produced by the pQE80L construct can go without this component.

To complete the dialysis, the protein solution was exposed to the first liter of solution placed in a 1000 mL graduated cylinder for a period of ninety minutes at 4 °C with moderate stirring. Following the first step the dialysis tubing was placed in the second liter of solution for a period of four hours at 4 °C , again with moderate stirring. The third step of the dialysis was left overnight under the same conditions in a third solution.

3.8 Concentration Procedures

The resulting solution from the dialysis was concentrated in an Amicon Ultra centrifugal filter device with a 30,000 molecular weight cut off. A Sorvall Legend RT table top centrifuge was used at 3150 rpm in ten minute intervals until the concentration was complete. The filtrate was discarded, while the concentrated solution was always poured back into the mother liquor to ensure the protein concentration within the centrifuge device was not increasing to precipitating levels.

3.9 Bradford Assay Analysis to Determine the Extinction Coefficient (ϵ) of MenD

In order to ensure a fast and repeatable measurement of the numerical value of protein concentration an experiment to link the Bradford Assay to absorbance at 280 nm (A_{280}) was performed. This was also done since it was found that the Bradford Assay used a larger volume of the purified and concentrated protein than the absorbance test would require.

Beer's Law, $A_{280} = \epsilon bc$, was utilized to determine the extinction coefficient (ϵ), where $b =$ the path length of light through the sample, and $c =$ concentration in units of mol/L (M). The protein concentration was found using a Bradford Assay. The pathlength (b) was the standard one centimeter. A reading of A_{280} was taken by first performing a blank/zero reading of one milliliter of pH 7.8 Tris-HCl buffer and then in a separate cuvette placing 10 μ L of protein sample from the same batch of protein utilized for the Bradford Assay into 990 μ L of pH 7.8 Tris-HCl buffer. From this information the extinction coefficient could be determined. Due to errors that commonly arise from the Bradford Assay the experiment was repeated ten-fold and utilizing a variety of protein concentrations to ensure that variable concentrations did not affect the result.

It was found that using 10 μ L of protein solution and diluting it as described above that an A_{280} reading of 0.0983 translated to 10 mg/mL of protein. Therefore, any protein sample could be analyzed using the following simple formula where C_{standard} is equal to 10 mg/mL and $A_{280, \text{standard}}$ is equal to 0.0983.

$$C_{\text{protein}} = A_{280, \text{protein}} * C_{\text{standard}} / A_{280, \text{standard}}$$

3.10 Crystallization Procedures

3.10.1 Sitting Drop Screening

Sitting drop crystallization experiments were set up at both room temperature and within a cold room (4 °C) using Nextal crystal screens supplied by Qiagen. The screens utilized for analysis include the Classics, Classics Lite, Cryos, Classics II, pH clear, and pH clear II, JCSG+,

PACT, Anions, Cations, PEG, PEG II, AmSO₄, MPD and ComPAS. All of these were set up in 96-well sitting drop trays using 100 µL of well solution, which was provided by the Nextal kits, and then utilizing 1.0 µL of well solution and 1.0 µL of protein solution. In some cases a 2:2 ratio was utilized to provide a slight alteration to the surface area of the drop. The wells are covered with a transparent tape as quickly as possible to ensure minimal evaporation to the environment.

3.10.2 Microbatch Screening

In the case of microbatch screening the Nextal screens provided by Qiagen were again utilized as was the case in the sitting drop experiments. Crystal plates were again set up in both the cold room (4 °C) and at room temperature (22 – 25 °C). However, in the case of this experiment seventy-two well microbatch plates were utilized. Therefore, it requires the use of two microbatch plates to do a single Nextal screen. In this experiment 1.0 µL of protein solution was mixed with 1.0 µL of well solution directly from the Nextal screen. The wells are then covered with Al's oil (50% paraffin oil and 50% silicon oil) to ensure minimal evaporation to the environment. The mechanism of crystallization differs when comparing that of microbatch with sitting drop because the latter relies on active diffusion whereas microbatch simply relies on mixing.

3.10.3 Hanging Drop and Optimization Methods

Crystal conditions that were found to produce crystals were optimized by creating a custom screen around that condition. For instance, if crystals were found in a well containing 0.1 M sodium acetate pH 5.0 and 40% MPD as well as in a well containing 0.1 M sodium acetate pH

6.0 and 30% MPD. An optimization could be set up to test several conditions that vary the pH of the sodium acetate between 5.0 and 6.0 and the concentration of MPD between 30% and 40%.

The most used procedure for optimization was to utilize both hanging and sitting drop. The hanging drop method utilizes the same concept but is set up slightly differently and often has a larger volume of protein and precipitant mixed together. 3.0 μL or 5.0 μL of each the protein sample and precipitant were utilized rather than the 1:1 or 2:2 ratio that was often used with the sitting drop technique. With the larger volume often large crystals were formed using this technique. That being said, the size of the crystal does not determine how well it diffracts the incoming x-rays.

3.11 Diffraction Procedures

Many of the crystals found within the screens and produced from optimization were tested for diffraction on the DX8 Proteum diffractometer at the Saskatchewan Structural Science Center at the University of Saskatchewan. Crystals were mounted on a loop and centered in front of the X-ray beam. Exposure times for diffraction were set at sixty seconds, the crystal was then rotated ninety degrees and exposed for an additional sixty seconds. For crystals that had some indication of a diffraction signal they were exposed a second time for a five minute interval at the angle that produced better results.

In all diffraction experiments, the crystals were not placed in a cryoprotectant since the exposure to X-ray beam was only one minute for each crystal. It is possible yet unlikely that the integrity of the crystals was compromised but all diffraction experiments were conducted with the same methods.

Chapter 4: Results and Discussion

4.1 Computational Work

In August of 2005 two high resolution diffraction data sets of crystals from the pMD14 construct of the enzyme were gathered by Edyta Sieminska. The two data sets were from an apoenzyme crystal and a holoenzyme crystal which was co-crystallized with both MenD cofactors; Mn^{2+} and ThDP. The first work that was done by myself was the processing of the acquired data and the following attempts to solve the phase problem using this data set for molecular replacement (MR). Many different software packages were used for this analysis and results of each will be discussed briefly.

4.1.1 MOSFLM and SCALA

The processing of the data sets was completed via the integration and scaling of the information provided by the diffraction images collected. This is done to determine the resolution limits of the images, the crystal mosaicity, the space group of the unit cell, the number of rejected reflections and the signal-to-noise ratio (I/σ) by sequentially using two separate software packages, MOSFLM (Leslie 1992) and SCALA (Weiss 2001).

With the use of MOSFLM, the crystal mosaicity of the holoenzyme was found to be 0.42 and the space group to be $P2_12_12_1$ with the following dimensions: $a = 106.686$, $b = 143.1776$, $c = 156.9498$. Utilizing SCALA to analyze the integrated data output from MOSFLM, it was found that the holoenzyme dataset was 97.3% complete at a resolution limit of 1.9 Å and a signal to noise ratio greater than 2.0 (Table 4.1A). The apoenzyme data indicated the same space group

with similar dimensions ($a = 106.4761$, $b = 143.0784$, $c = 159.037$) but had a higher mosaicity value of 0.51. The data for the apoenzyme was also only 68.3% complete at a resolution limit of 2.4 Å, and indicating a signal to noise ratio greater than 2.0 (Table 4.1B).

Table 4.1 - Output data from SCALA to determine the resolution limit. The first column of each table indicates the series of values of increasing resolution. The second column indicates a value of R_{merge} ; a value of 0.400 was utilized to select resolution limit with R_{merge} . The third column indicates the signal to noise ratio (I/σ); a value of 2.0 was utilized to select resolution limit with I/σ . Table A shows the holoenzyme data; Table B shows the apoenzyme data

A			B		
D_{min} (resolution (Å))	R_{mrg}	I/σ	D_{min} (resolution (Å))	R_{mrg}	I/σ
5.06	0.021	26.5	5.55	0.024	21.1
3.58	0.027	18	3.93	0.032	18.6
2.92	0.036	16.1	3.21	0.052	13.3
2.53	0.059	12.6	2.78	0.123	6.1
2.26	0.091	8.1	2.48	0.302	2.4
2.07	0.193	3.8	2.27	0.619	1.2
1.91	0.353	2.7	2.10	1.51	0.5
1.79	0.704	1.6	1.96	2.675	0.3
1.69	1.143	0.9	1.85	7.813	0.1

The output from SCALA is a reflection data file which is an encoded file containing the intensity of reflections. This file is developed with the specific purpose of being input for further analysis using a variety of other software programs used to solve the phase problem. The data for the holoenzyme was selected to be used for further analysis with molecular replacement programs since it had a higher resolution, higher percent completeness and less crystal mosaicity than the apoenzyme.

4.1.2 MRBUMP

The first program that was attempted for molecular replacement was a novel software at the time called Molecular Replacement By Using Multiple Proteins (MRBUMP, Keegan 2006). MRBUMP requires nothing but a protein sequence and a reflection data file to perform a sophisticated and automated search for the best molecular replacement solution. It does this by performing a sequence alignment with all known proteins in the RCSB protein databank and then selecting those with the best alignment scores for further processing. The selected molecules then undergo molecular replacement and, if prompted to, a determined number of rounds of refinement.

Utilizing MRBUMP to find a molecular replacement solution for MenD resulted in no hit protein sequences within the protein data bank. In 2007, many alterations were made to the MRBUMP software suite that allowed the user to alter the minimum sequence identity the program accepts. Unfortunately, even with the improvements to this programs interface and the ability to alter the program parameters, a suitable molecular replacement solution is still not provided by MRBUMP for MenD. The most common result of the usage of MRBUMP in this project was an error message that indicated that no structure within the PDB was sufficiently similar for the program to proceed.

4.1.3 AMoRe

The second program that was utilized for molecular replacement was AMoRe (A Molecular Replacement program, Navaza 1994). This software performs a rotation and a translation function of the input model into the signal designated by the input reflection data file.

AMoRe then outputs how well the input model correlates to the signal present to give a correlation coefficient and an R_f value. An ideal solution will be one that has a high correlation coefficient and the lowest R_f value.

R_f is a measure of model quality comparing $|F_{obs}|$ and $|F_{calc}|$:

$$R_f = \frac{\sum | |F_{obs}| - |F_{calc}| |}{\sum |F_{obs}|}$$

The measured structure-factor amplitude, ($|F_{obs}|$), is calculated from reflection intensities from the data set and $|F_{calc}|$ is the calculated amplitude from the model provided. An R_f value of 0.60 (often displayed as 60 in AMoRe (as a percentage value)) indicates that the measured amplitudes are being compared to random amplitudes (Rhodes 2000).

The correlation coefficient is an alternative mathematical comparison of the amplitude of the observed structure factor within the data and the calculated structure factors provided by the model. The scale used for the correlation coefficient within AMoRe is a percentage value from 0-100 (Navaza 1994).

The reflection data file produced by SCALA was utilized within AMoRe by comparing the data to seventeen enzyme coordinate files from the protein data bank to produce a molecular replacement solution. Along with these computational experiments being completed, polyalanine co-ordinate files of the same seventeen models were generated using MOLEMAN2 (Kleywegt 1998) and analyzed with AMoRe. The best result produced by AMoRe was with the model of yeast AHAS bound to an inhibitor tribenuron methyl (PDB code, 1T9A). For this model AMoRe

gave a peak R_f value of 54.3 and a second best R_f value of 55.1. The correlation coefficient (CC) showed a maximum value of 8.9 and a second best value of 8.4. These results were deemed the best because of the presence of a gap between the best and second best result for both the values of R_f and the correlation coefficient. The variance in top results for R_f and CC were not present in any of the other models analyzed by AMoRe. Such a gap can often be indicative of a result that is not just random alignment. A complete chart of the results obtained from all thirty-four models is seen in Table 4.2. None of these results were improved with stepwise refinement using Refmac5 (Murshudov 1997).

The results produced using AMoRe were deemed insufficient for a structural solution and the software package was not used again for the remainder of the research project.

4.1.4 PHASER

Although molecular replacement was unsuccessful using AMoRe, another software suite known as PHASER (Read 2001) was used to test twenty-five PDB models of interest, which also now included the models for benzaldehyde lyase ((BAL), 2AG0 and 2AG1). The structure of this enzyme was published (Mosbacher 2005) during the analysis with PHASER. PHASER utilizes more complex algorithms than AMoRe and also includes an iterative refinement process.

The input for PHASER is much the same as for AMoRe, but the output is not in the form of R_f values and correlation coefficients but rather as Z-scores and log likelihood gain. Z-scores are defined as the likelihood that the result is correct. The magnitude of the Z-score is determined by the number of standard deviations the result is away from the mean of all results. A Z-score of eight or greater is said to indicate that phaser worked as a tool for MR, whereas a

Table 4.2 – Output results from AMoRe. The first column indicates the protein data bank model utilized for analysis. The second column provides the sequence identity of the protein in comparison to the MenD protein sequence. CC (2nd) refers to the best and second best result produced for correlation coefficient as produced by AMoRe. R_f (2nd) refers to the best and second best result for R_f values as produced by AMoRe. The PDB form indicates whether the coordinate file was for a dimer or tetramer. The following two columns are for the polyalanine models. The final column indicates the protein that is being analyzed.

Model Used	Seq. Identity	CC (2nd)	Rf (2nd)	PDB form	CC_Pala (2nd)	Rf_Pala (2nd)	Protein
1PYD	-	9.4 (9.3)	55.1 (55.0)	tetramer	9.3 (9.2)	55.4 (55.4)	Catalytic Centers of PDC
1ZPD	27.273	7.9 (7.9)	54.4 (54.4)	tetramer	13.2 (12.0)	58.0 (61.8)	PDC from <i>Z. Mobilis</i>
1N0H	19.361	7.3 (7.3)	54.6 (54.7)	dimer	8.3 (8.2)	54.8 (54.9)	AHAS bound to Chlorimuron
1T9A	19.361	8.9 (8.4)	54.3 (55.1)	dimer	8.6 (8.5)	54.7 (54.9)	AHAS bound to Tribenuron Methyl
1T9B	19.361	8.2 (8.2)	55.2 (54.6)	dimer	8.2 (8.1)	55.3 (55.7)	AHAS bound to chlorsulfuron
1T9C	19.361	8.4 (8.4)	54.6 (54.6)	dimer	8.4 (8.3)	55.0 (54.8)	AHAS bound to Metsulfuron Methyl
1T9D	19.361	8.3 (8.0)	54.3 (54.5)	dimer	7.7 (7.7)	54.5 (54.6)	AHAS bound to Sulfometuron
1JSC	19.361	9.0 (8.9)	54.4 (54.6)	dimer	8.4 (8.4)	54.7 (54.8)	AHAS from <i>S. Cerevesiae</i>
1OZG	18.891	8.8 (8.6)	54.3 (54.5)	tetramer	8.3 (8.3)	54.6 (54.6)	ALS tetramer with unusual intermediate
1OZH	18.891	8.4 (8.3)	54.4 (54.6)	dimer	8.3 (7.8)	56.9 (56.6)	ALS dimer with unusual intermediate bound
1OZF	18.891	8.5 (8.3)	54.6 (54.5)	tetramer	8.0 (8.0)	54.7 (54.8)	ALS from <i>K. Pneumoniae</i>
1PO7	20.074	8.7 (8.6)	54.5 (54.7)	tetramer	8.7 (8.6)	54.5 (54.6)	BFD E28A mutant
1Q6Z	20.074	8.6 (8.6)	54.4 (54.7)	tetramer	8.5 (8.5)	54.5 (54.7)	BFD E28A mutant with thiamine thiazoloone diphosphate bound
1PI3	20.074	9.0 (8.8)	54.5 (54.5)	tetramer	8.7 (8.7)	54.3 (54.5)	E28Q mutant of BFD
1BFD	20.074	9.0 (8.9)	54.5 (54.8)	tetramer	8.8 (8.7)	54.6 (54.8)	BFD from <i>P. Putida</i>
1MCZ	20.074	error	error	tetramer	Error	Error	BFD bound to <i>R</i> -mandelate
1OVM	29.167	9.2 (9.1)	54.6 (54.6)	tetramer	9.0 (9.0)	54.6 (54.8)	IPDC from <i>E. Cloacae</i>

score below five indicates that it definitely did not work; intermediate values are less conclusive. LL-gain measures the difference of likelihood that the data can be predicted to match with your model than with a random distribution of the same atoms (Read 2001).

Of the twenty-four models tested the best result was found with PHASER was the polyalanine model of benzaldehyde lyase (pAla-2AG0), which showed a Z-score of 10.34 and a LL-gain of +79.50. This was the only PDB file that produced a Z-score that exceeded eight (complete data in Table 4.2). Initial analysis using Refmac5 (Murshudov 1997) showed that the model produced had an R_f value of 52.1 and with stepwise refinement this was brought down to 51.5 (refined model shown in Figure 4.1).

However, despite continuous effort over several months to alter the alanine residues to that of the MenD sequence, and not seeing any positive results in the way of dropping R_f values, it was determined that the result was a false positive. It was shown to be highly model biased and was basically a duplicate of the benzaldehyde lyase polyalanine model. This was tested via a least squares superimposition using LSQMAN with the PHASER molecular replacement result and the polyalanine benzaldehyde lyase. The result of this analysis was a root mean squared deviation (RMSD) of less than 0.01 Å. This indicates that benzaldehyde lyase and the pAla model that was produced by PHASER was essentially an exact replicate of benzaldehyde lyase.

The results that were produced by PHASER encompassed several months of time to produce. The results led us to believe that the structure could and would be solved using molecular replacement strategies. However, it was found eventually that no means were available to solve the structure using PHASER. As such, other methods needed to be utilized.

Table 4.3 - Output results from PHASER. The first two columns of data indicate the output Z-score and the value of log-likelihood gain following the rotation function. The next two columns displays the output Z-score and the value of log-likelihood gain following the translation function. The LL-gain refined indicates the log-likelihood gain following anisotropic corrections.

PDB Search file	Z-Score (RF)	LL-gain (RF)	Z-Score (TF)	LL-Gain (TF)	LL-gain refined
Benzaldehyde Lyase (BAL)					
2AG0 (tetramer)	5.042	2.21	6.48	-154.05	-149.66
2AG0 (monomer)	4.524	20.61	5.79	32.79	36.90
2AG0 (polyAla)	5.829	40.17	10.04	76.31	79.50
2AG1	4.994	2.81	6.68	-133.12	-145.80
2AG1(polyAla)	4.913	30.90	7.41	25.68	40.78
Benzoylformate Decarboxylase (BFD)					
1BFD (mono)	4.458	19.91	5.95	32.64	34.47
1BFD (polyAla)	4.555	18.97	6.01	21.52	24.57
Acetolactate Synthase (ALS)					
1OZF (dimer)	4.197	25.05	5.85	24.67	28.87
1OZF (monomer)	4.939	21.89	5.84	32.72	38.41
1OZF (polyAla)	5.155	30.81	5.28	42.84	43.33
Pyruvate Decarboxylase from <i>S. cerevisiae</i> (ScPDC)					
1QPB (dimer)	4.556	23.63	6.00	24.58	22.08
1PVD (dimer)	5.082	29.75	5.13	24.23	32.81
1PVD (poly Ala)	4.987	27.21	5.07	29.83	33.45
Pyruvate Decarboxylase <i>Z. mobilis</i> (ZmPDC)					
1PYD (dimer)	5.414	35.99	5.56	2.76	9.42
1ZPD (tetramer)	4.418	-54.75	5.64	-397.11	x
1ZPD (monomer)	4.787	23.12	5.78	35.57	40.65
1ZPD (polyAla)	5.419	28.55	5.40	-73.26	-62.87
Indolepyruvate Decarboxylase (IPDC)					
1OVM (tetramer)	4.754	-48.12	5.40	-425.96	-416.92
1OVM (monomer)	4.693	28.04	5.81	36.28	43.35
1OVM (polyAla)	5.025	25.73	5.47	-45.88	-41.85
Acetohydroxy Acid Synthase (AHAS)					
1JSC (dimer)	5.069	35.08	6.03	39.44	41.24
1JSC (monomer)	4.765	23.50	5.83	41.58	41.96
1JSC (polyAla)	4.762	10.87	5.01	20.74	23.19
1T9A (dimer)	5.060	25.40	6.39	20.31	28.49
Theoretical Model					
1XV2	4.167	17.60	x	x	x

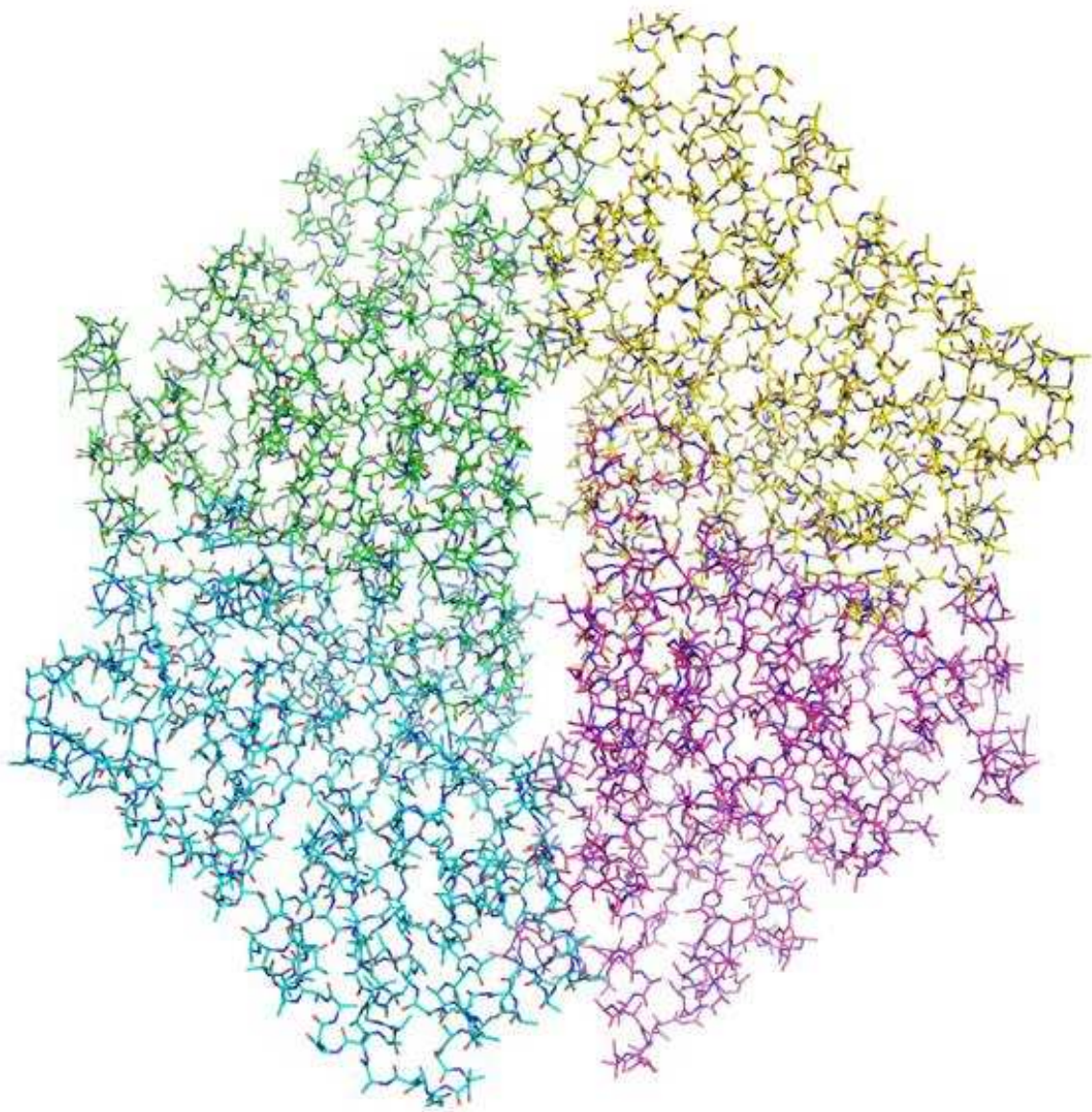


Figure 4.1 - PDB coordinate file produced by PHASER using 2AG0 polyaniline tetramer, Z-score = 10.34, LL-gain (refined) = 79.50.

4.1.5 Modeller

Since molecular replacement proved to be unsuccessful with models provided by the PDB it was thought that the generation of a model using the program *Modeller* (Fiser 2000)

could provide us with some insights into the orientation of the MenD active site and also provide an additional model to be tested for molecular replacement using PHASER.

Modeller requires the user to provide an input PDB file and an input sequence alignment file. The sequence alignment file is composed of two proteins; the desired protein and that of the sequence of the associated input PDB. Based on the alignment, the structural implications of sequence, and the spatial restraints, a model of best fit containing the MenD sequence and the commonly observed ThDP dependent enzyme fold was generated. Given the results produced with PHASER, where a benzaldehyde lyase polyalanine model produced a Z-score greater than eight, it was presumed that the backbone structure of MenD and BAL were similar. As such, benzaldehyde lyase was utilized as the template for the MenD *Modeller* model.

This generated model was then used as a search model for MR. This as well proved to be unsuccessful as the Z-score in PHASER was 5.02 and showed no gap between the top ten Z-scores shown as results. As such the model (Figure 4.2) proved incapable of solving the phase problem.

With the phase problem not able to be solved via molecular replacement, the *Modeller* model was analyzed in an attempt to gain understanding of MenD chemistry from the residues present within the active site and ThDP binding site (discussion in section 4.3). Binding of ThDP and substrates to ThDP-dependent enzymes requires a dimeric structure and the model produced by *Modeller* was only a monomer. Thusly, the program LSQMAN was utilized to produce a dimer by superimposing the *Modeller* model onto the A chain and C chain of benzaldehyde lyase to produce a tight dimer (Figure 4.3A). The generated AC MenD dimer model was then utilized

to iteratively produce a tetramer by superimposing it on the BD chain dimer of benzaldehyde lyase (Figure 4.3B).

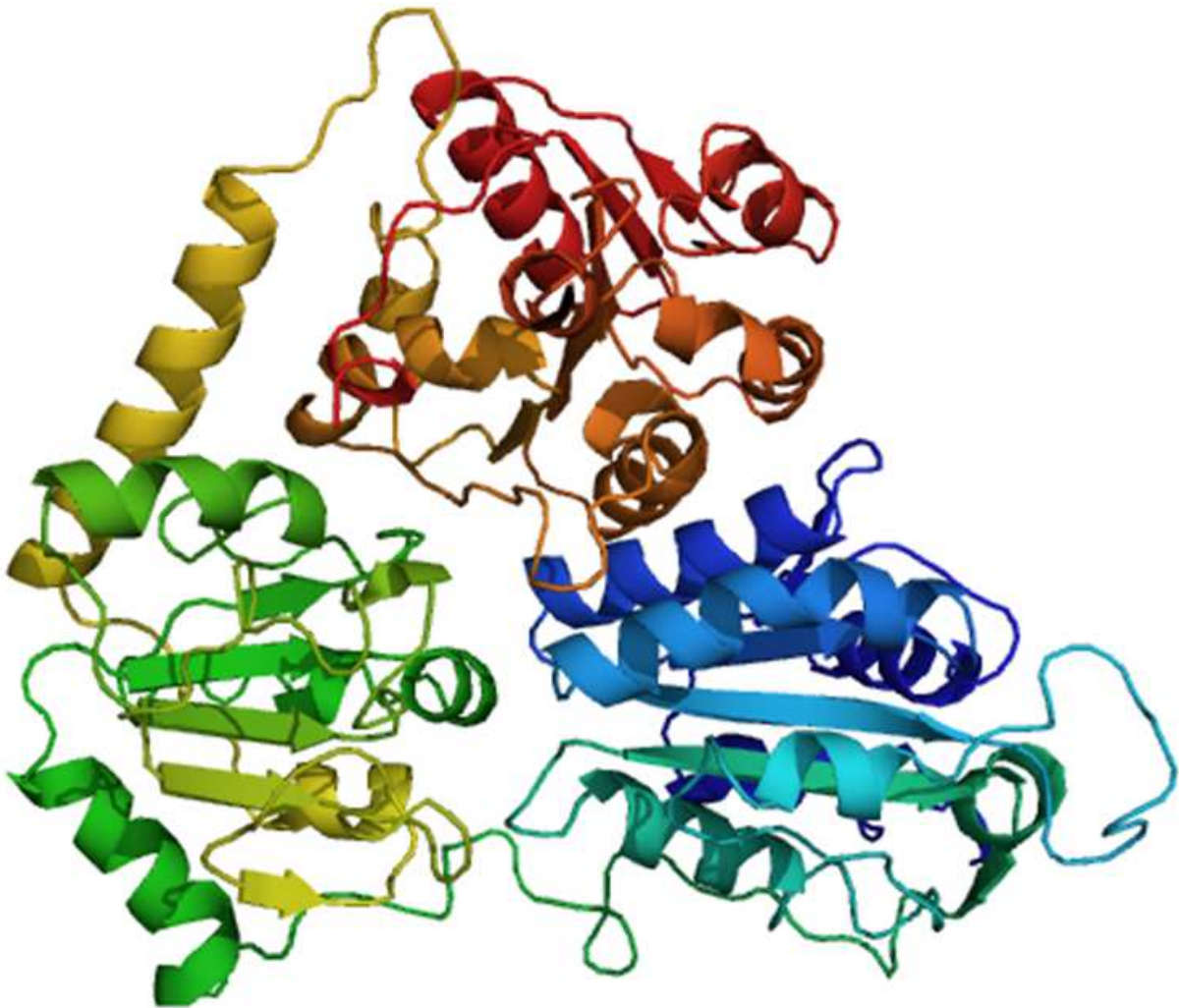


Figure 4.2 - Ribbon Diagram of monomer model produced by *Modeller* containing the MenD sequence; viewed in PyMol

LSQMAN was also utilized to do a least square analysis as well to see how similar the *Modeller* model is to that of the PHASER polyaniline output result (Figure 4.1) and also to that of benzaldehyde lyase (Figure 1.14). The result of this analysis indicated an RMSD of 0.10 Å for both models. An RMSD of this magnitude shows that the models are almost identical. Also,

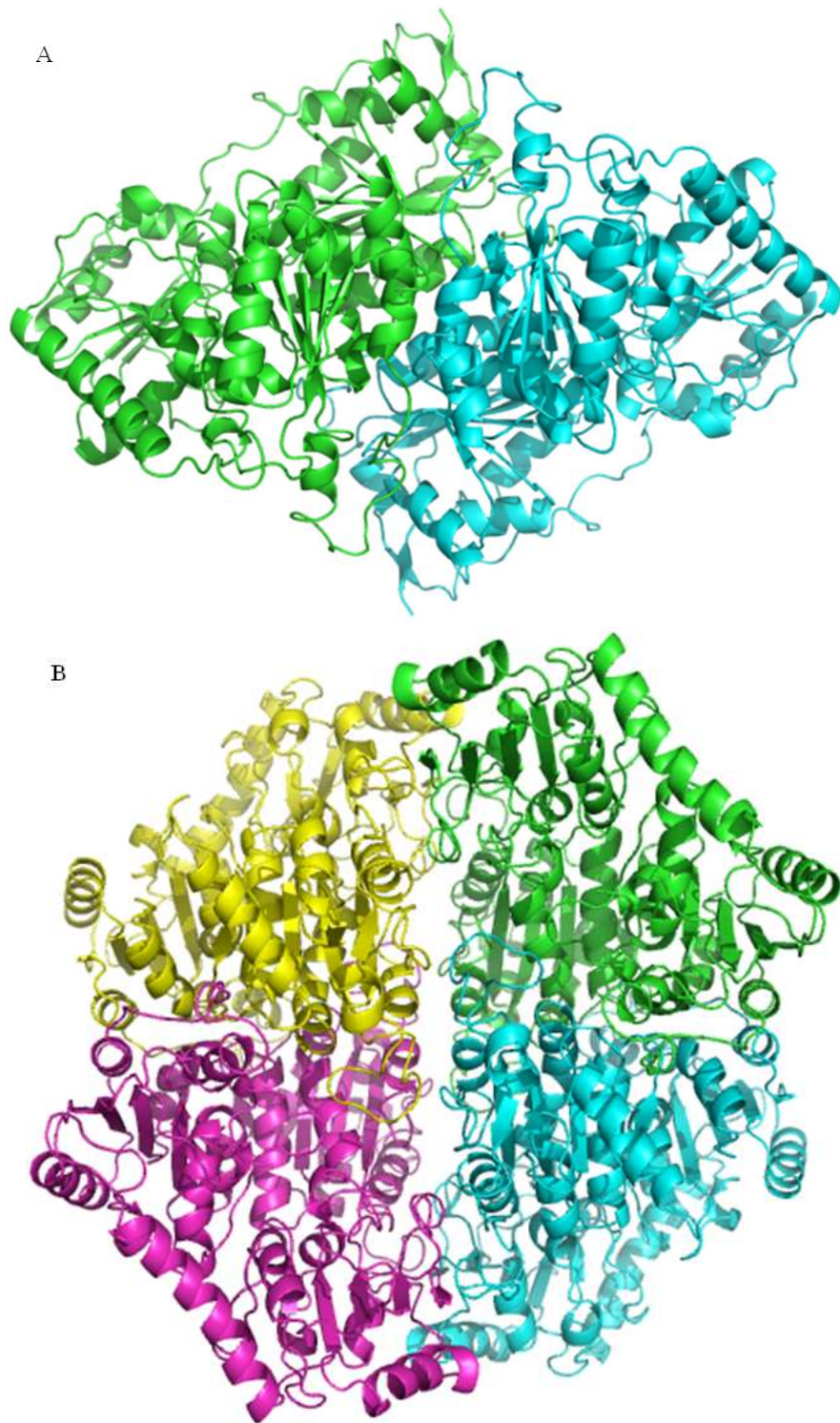


Figure 4.3 - A) generated dimer from the *Modeller* model B) generated tetramer from the *Modeller* model

alignment using PyMol shows that the supersecondary structure is almost identical and only the side chains differ when comparing benzaldehyde lyase (2AG0) to the *Modeller* model (Figure 4.4). This indicates that no substantial spatial constraints were faced when generating the model with the MenD sequence in place of the BAL sequence.

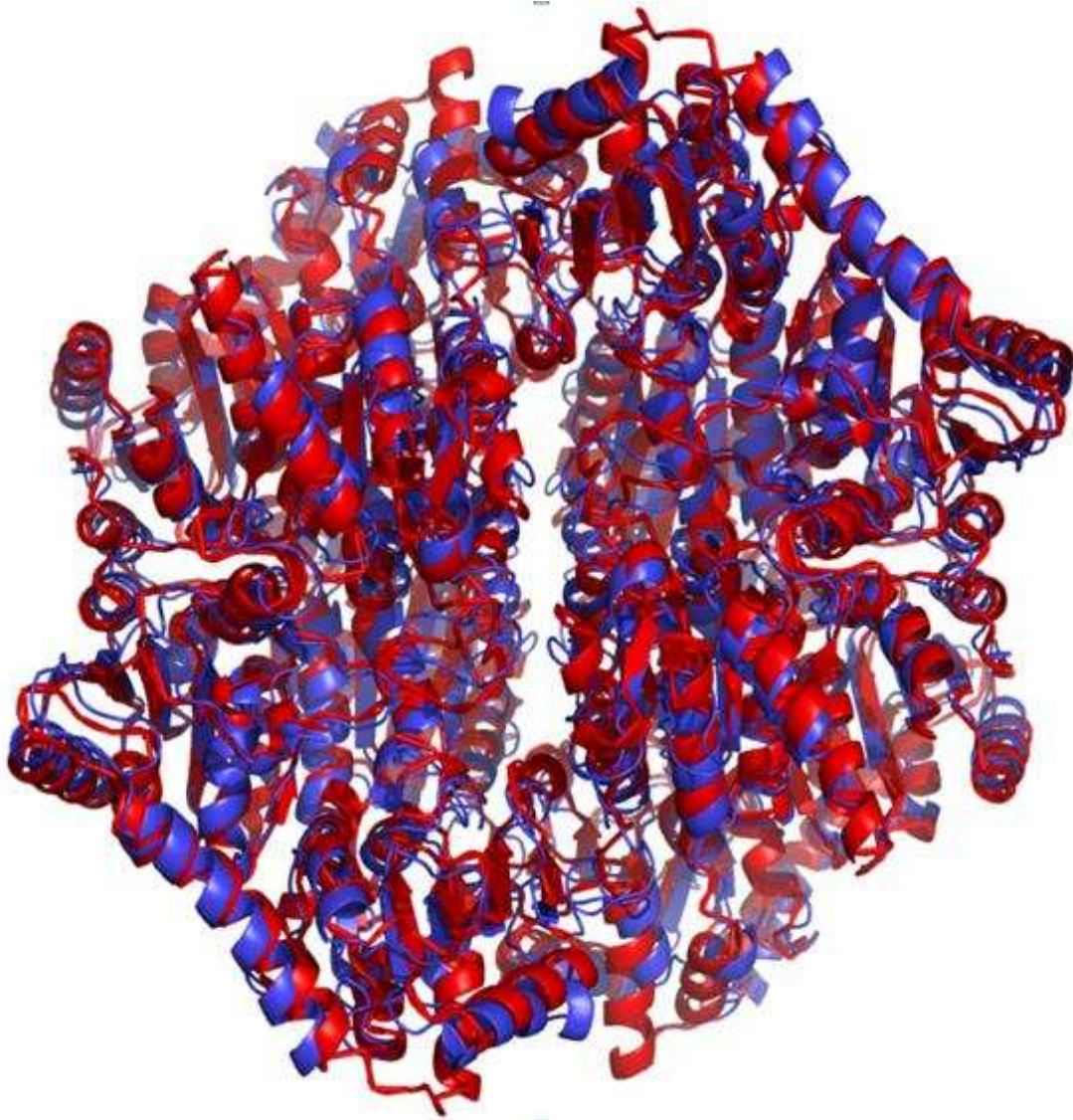


Figure 4.4 - Superimposition of MenD *Modeller* model tetramer (blue) and the PDB file produced by PHASER (red) in molecular replacement. Only notable difference is seen in the random loop of the middle domain seen at the top and bottom of the figure in red. LSQMAN reports RMSD = 0.10 Å

4.2 Expression of the MenD Gene

4.2.1 Expression of MenD from pMD14b in BL21

Previous research done in the lab by Edyta Sieminska provided glycerol cells stocks of BL21 cells containing the pMD14b plasmid. Also, purified pMD14b plasmid that could be transformed into competent cells was also stored at -80 °C. However, upon attempting to express this gene from both the glycerol stocks of BL21's and newly transformed BL21, Rosetta and DH5 α cells, no major over-expression band was observed on an SDS-PAGE gel. (Figure 4.5A) Eventually, a transformed BL21 cell batch produced MenD protein (Figure 4.5B), but the concentration of MenD protein obtained within the soluble fraction was not sufficient for crystal screening to be done. The amount of protein that was expressed in the soluble fraction was less than 1 mg/L of media and other approaches for producing soluble MenD protein were investigated.

4.2.2 Selection of a New Vector

The phase problem could not be solved with the 1.9 Å wild type data set collected and it was believed that further developments with the project could not be made with wild-type crystal diffraction. Therefore, a heavy atom derivative was desired in order to perform multiple wavelength anomalous dispersion (MAD) to solve the phase problem. The vector selected would be one that expressed SeMet proteins with high reliability.

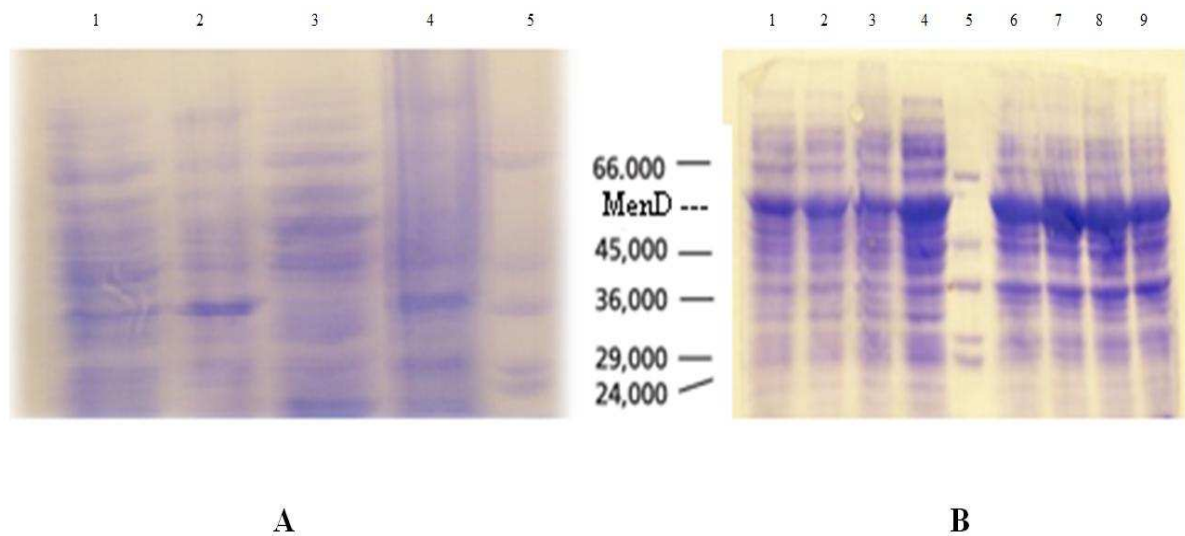


Figure 4.5 - SDS-PAGE gel showing: A) Inadequate production of MenD protein in: 1A BL21 glycerol stocks; 2A BL21; 3A Rosetta; 4A DH5α; marker B) Production of MenD protein from BL21 cells, 1-4B before lysis; 5B marker; 6-9B after lysis with lysozyme. The molecular weight for MenD is expected to be 61.4 KDa.

A specific type of Qiagen vector known as pQE80L (Qiagen, Figure 4.6) was selected as the new target plasmid for the MenD gene to be cloned into. This plasmid provides various useful elements that were also present on pMD14b, including the same antibiotic selectable marker and an N-terminal hexa-histidine tag. The pQE80L plasmid contains a pair of lac repressors, making the expression of the protein more controlled, since expression of the target gene becomes completely dependent on the induction of the cells with IPTG (Qiagen 2003). Also, pQE80L can be expressed in a cloning strain and does not require purification and transformation of the plasmid into an expression strain such as BL21.

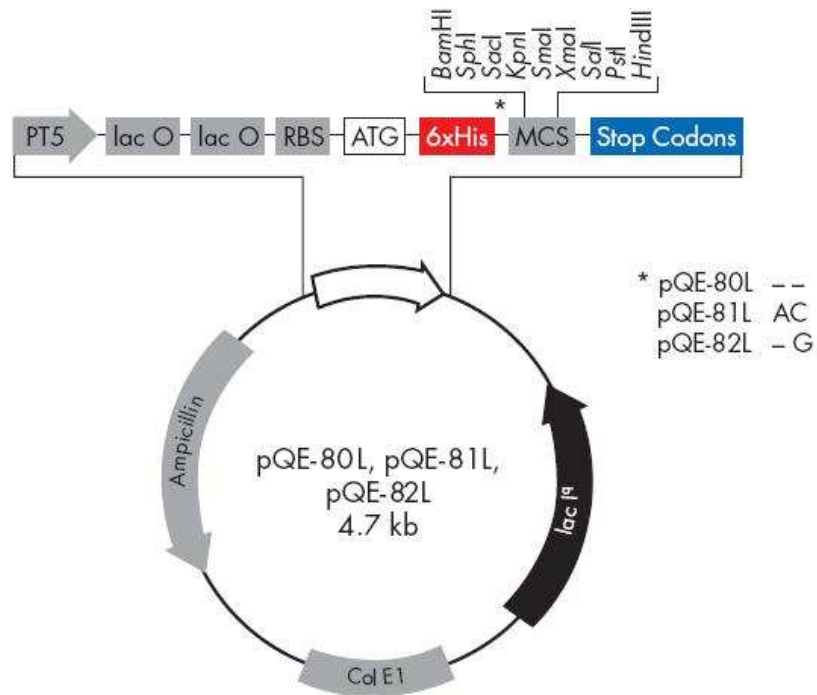


Figure 4.6 - Map of pQE80L as provided by Qiagen (Qiagen 2003)

It was also thought that the use of pQE80L would allow for one of the large problems faced previously in the preparation of SeMet MenD protein to be overcome. Use of pMD14b resulted in leaky expression, meaning that protein was being expressed prior to the induction with IPTG and therefore wild-type (non-SeMet) protein was produced. It was thought that the utilization of pQE80L will allow for complete inclusion of SeMet into the protein produced.

4.2.3 Cloning of MenD gene into pQE80L

Primers capable of amplifying the MenD from purified pMD14b and also mutating the sequence such that a BamHI site was present upstream of the start codon and also a KpnI restriction site downstream of the stop codon were designed. This was done such that upon restriction digest of PCR product (containing the amplified MenD gene) and pQE80L with KpnI and BamHI (Figure 4.7), a subsequent ligation would result in the formation of the new plasmid

construct. Although several protocols were attempted, positive results were observed with the following experimental conditions: PCR amplification (annealing temperature 63 °C and annealing time 60 seconds, 30 cycles); restriction digestion (37 °C, 3 hours); ligation (25 °C (desktop/room temperature) 3 hours); transformation into Nova Blue cells via the heat shock method; and subsequent plating onto LB/agar plates. It may be noteworthy that the ligation required a far greater ratio of cut amplified product to cut plasmid than the suggested 3:1 ratio. Approximately one-hundred colonies were observed when the transformation was done on a 10:1 ratio ligation mixture, but as the ratio was decreased fewer colonies were observed.

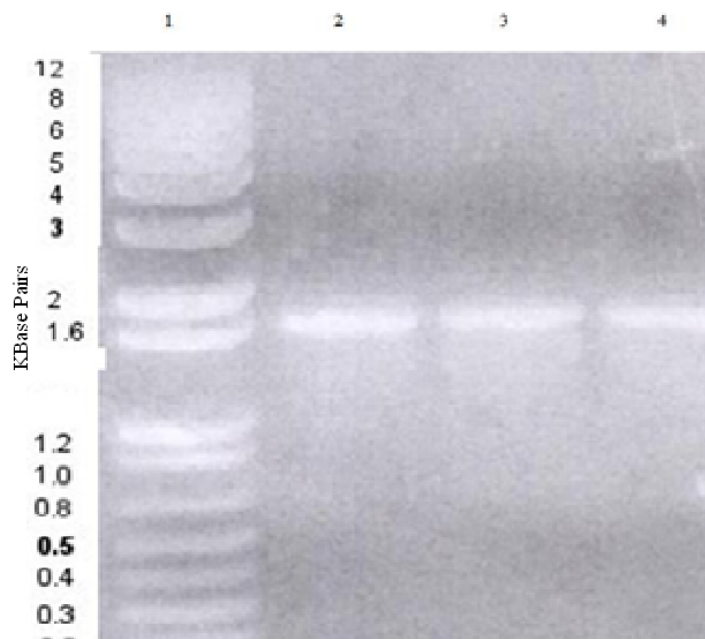


Figure 4.7 - 10% Agarose gel with 1Kb Plus Marker (lane 1, Kb pairs shows beside) and PCR amplified MenD gene following a) single digest with BamHI, b) single digest with KpnI, c) double digest. The experiment shows that the target gene (1.7 Kb pairs) was cut as intended and did not contain internal KpnI or BamHI sites as was expected

The resultant colonies that grew on the plates could have been pQE80L that had re-annealed to itself or pQE80L fused with our cut DNA insert of interest. Since both would confer ampicillin resistance, both could allow colonies to grow on the plate. As such, a colony PCR experiment was performed on each single colony formed on the plates. The colony PCR experiment utilized an almost identical procedure to that of the initial PCR experiment. Positive inclusion of the MenD gene into the pQE80L plasmid was indicated with a 1700 base pair band visualized using agarose gel electrophoresis. The band size can be measured by being compared to 1 Kb Plus marker.

4.2.4 Expression of MenD from pQE80L Construct

MenD protein was expressed from the new construct within Nova Blue cells. This was accomplished by inducing the cell grown in 1.0 L of LB with 0.4 mM IPTG at OD₆₀₀ equal to 0.4. A protein band viewed on an SDS-PAGE gel (Figure 4.8, lane 4) confirmed the over-expressed protein shared similar size to that seen with pMD14b expressed protein (Figure 4.5). Nearly twenty milligrams of over-expressed protein were recovered following purification, dialysis and concentration of protein. Resultant SDS-PAGE gel following expression and purification is seen in Figure 4.8 (lane 2 and 3). The purification, dialysis and concentration process was the same that was followed with pMD14b plasmid expressed protein as described in materials and methods.

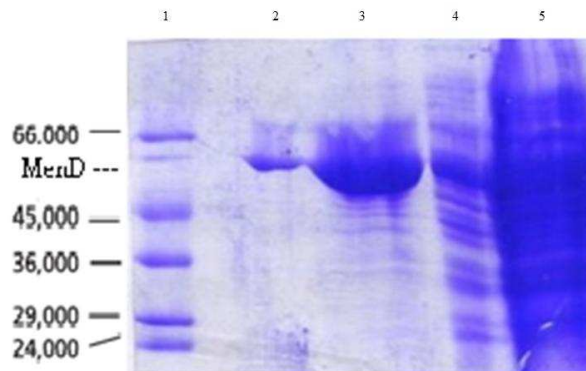


Figure 4.8 - SDS-PAGE Gel: Lane 1 – marker, Lane 2 - 3 μ L of concentrated and purified protein, Lane 3 – 10 μ L of concentrated and purified protein, Lane 4 – unpurified soluble fraction, Lane 5 – insoluble pellet protein

4.2.5 Testing Purified Protein for MenD Activity

The protein sample obtained was then tested for MenD reaction activity to prove the new construct was a functional MenD enzyme. This was done using a coupled assay which utilizes lactate dehydrogenase, NADH and pyruvate, which was previously believed to be a product of the MenD reaction. This assay allows for the progress of the MenD reaction to be monitored with decrease absorption of 340 nm light. NADH, which has an absorption peak at 340 nm, is reacted to produce NAD^+ as pyruvate is evolved and utilized as a substrate for lactate dehydrogenase (Figure 4.9). Protein expressed from the new construct (which was called pMD80L), indicated a V_{max} value of 0.97 $\mu\text{M}/\text{min}$ when an enzyme concentration of 90 nM was utilized. This result was comparable to that of the pMD14b enzyme of the same concentration ($1.0 \pm 0.3 \mu\text{M}/\text{min}$) (Ho 2007).

Additional research was done by Maohai Fang utilizing the new methods for MenD kinetic analysis with SEPHCHC as the product (Jiang 2007). MenD from pMD80L was found to have a k_{cat} value of $18 \pm 2 \text{ min}^{-1}$. This result was comparable to the k_{cat} value of $4.5 \pm 0.4 \text{ min}^{-1}$

published for MenD (Jiang 2007). These kinetic results confirm that cloning procedures were successful and the protein expressed from the gene cloned into pQE80L is MenD.

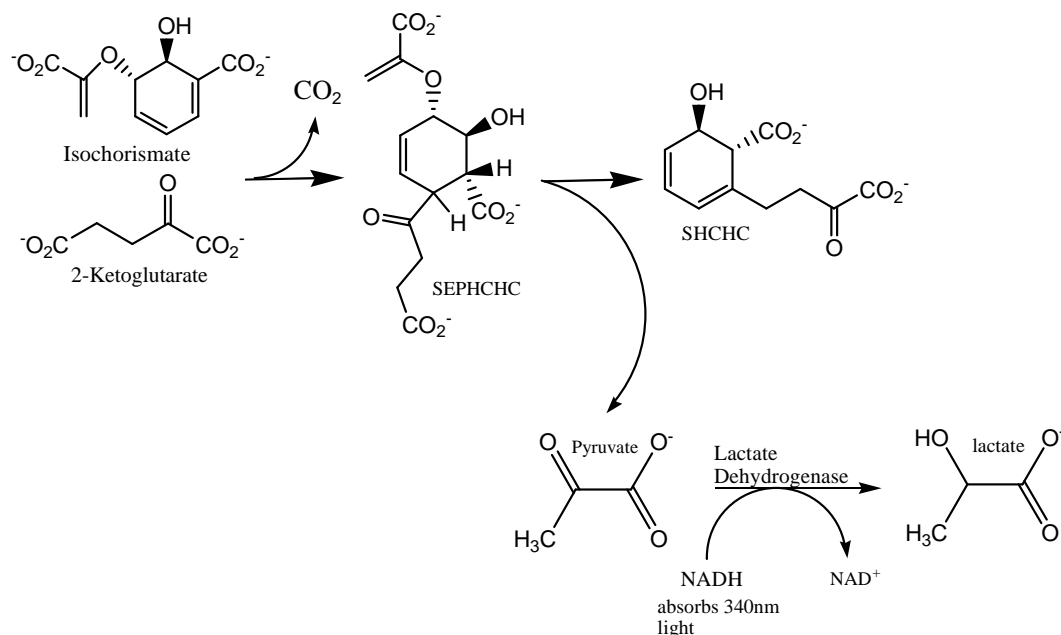


Figure 4.9 - Lactate dehydrogenase assay for monitoring MenD Activity. The spontaneous production of SHCHC from SEPHCHC produces pyruvate which is then converted to lactate by lactate dehydrogenase (LDH). The reaction with LDH results in the oxidation of NADH to NAD⁺ and thus causes a loss in absorption at 340nm within the reaction mixture. The rate of reaction is determined by the rate of decrease in absorption at 340nm.

4.2.6 Crystal Trials and diffraction of Native pMD80L protein

The Nextal crystal screen suites provided by Qiagen were used for initial screening. All screens (both microbatch and sitting drop technique) were tested with 20 mg/mL protein concentration. Analysis of the crystal plates showed that nearly seventy percent of the wells contained precipitate and no crystals were discovered. As such, the concentration was halved to 10 mg/mL. Upon setting up the screens with this concentration of protein a reduced number of precipitation events occurred but still no protein crystals were found.

Next, all contents of the dialysis buffer were removed except for the 50 mM Tris-HCl pH 7.8 and the cofactors ThDP and Mn^{2+} . The dialysis buffer utilized was designed for pMD14b construct to decrease occurrence of protein precipitation upon dialysis and concentration. It was found that removal of 12.5% glycerol and 200 mM NaCl had no effect on protein solubility with pMD80L. These two contents could potentially have hindered the crystallization process by acting like contaminants which prevented ideal protein-protein interactions.

Upon re-screening with the protein dialyzed (three-fold dialysis) into only 50 mM Tris-HCl pH 7.8 and the cofactors many conditions were found to produce crystals. Those conditions that produced crystals via the sitting drop method are listed in Table 4.4. All of the conditions listed were tested on the in-house Proteum diffractometer. The results of the diffraction experiments produced some results worthy of mention and are included in Table 4.4 as well. It became noticeable that the conditions containing a buffer (Tris / HEPES) with pH between seven and eight, and MPD with concentrations 30% to 50% weight- to-volume ratio, produced crystals that showed some diffraction. The listing of the result 'diffraction' is indicative of just a few protein spots present on the diffraction image with resolution being approximately 6-15 Å in most cases. An example of crystals that showed diffraction is shown in Figure 4.10.

Optimization of the HEPES buffer (pH 7.0, 7.5, 8.0, 8.5) and TRIS buffers (pH 7.0, 7.5, 8.0, 8.5) was done in varying concentrations of MPD (0%, 5%, 10%,...50%, 55% (w/v)). From this analysis it was found that far more crystals were obtained in the pH 7.5 HEPES buffer than in all of the other HEPES and Tris conditions combined. A second optimization screen was then done with 0.1 M HEPES buffer pH 7.5 that added PEG4000 to solution (2.5%, 5%, 7.5%, 10% PEG4000) to six different MPD conditions (25%, 27.5%, 30%, 32.5%, 35%) in the first forty-

eight wells, and the same conditions in the absence of PEG4000 in the other forty-eight wells. Crystals were produced overnight in 35% MPD, 7.5% PEG4000 and 0.1 M HEPES buffer pH 7.5 (Figure 4.11). It was also found with more testing that this result was reproducible and could also be done successfully via the hanging drop method as well. Hanging drop experiments were done with 5 μ L of protein solution and 5 μ L of well solution, leaving 995 μ L in the well. Results showed that the hanging drop method produced parallel results to that of the sitting drop method; however, larger crystals were produced with the hanging drop methods (Figure 4.12).

Table 4.4 - Crystallization conditions that produced crystals at [protein] of 10mg/mL and diffraction results from the in-house DX8 Proteum diffractometer.

Crystal Screen / well	Crystal condition	diffraction
ComPAS D1	0.1 M TRIS pH 8.5, 10 %(w/v) Ethanol	none
ComPAS E1	0.1 M HEPES pH 7.5, 5 %(w/v) Isopropanol	none
ComPAS E5	0.05 M Sodium chloride, 0.1 M TRIS pH 8.5, 12 %(w/v) Isopropanol	salt
ComPAS E6	0.2 M Sodium citrate, 0.1 M MES pH 6.5, 15 %(w/v) Isopropanol	salt
JCSG+ D1	24 %(w/v) PEG 1500, 20% (w/v) Glycerol	none
JCSG+ D5	0.1 M HEPES pH 7.5, 70 %(v/v) MPD	diffraction
JCSG+ D6	0.2 M Magnesium chloride, 0.1 M TRIS pH 8.5, 20 %(w/v) PEG 8000	none
JCSG+ D7	0.2 M Lithium sulfate, 0.1 M TRIS pH 8.5, 40 %(v/v) PEG 400	none
Cryos B3	0.018 M Calcium chloride, 0.09 M Sodium acetate pH 4.6, 27 %(v/v) MPD	salt
Classics C3	0.1 M TRIS.HCl pH 8.5, 2.0 M Ammonium phosphate	salt ring
Classics B10	0.1 M HEPES pH 7.5, 70 %(v/v) MPD	diffraction
Classics Lite C10	0.2M K/Na tartrate, 0.1M tri-Na citrate pH 5.6, 1M NH ₄ ⁺ sulfate	salt/ice ring
Classics2 C2	1.1 M Ammonium tartrate pH 7.0	salt/ice ring
MPD F10	0.1 M TRIS pH 8.0, 65 % (v/v) MPD	none
MPD F11	0.1 M TRIS pH 8.0, 65 % (v/v) MPD	none
MPD H4	0.1 M HEPES pH 7.5, 30 % (w/v) MPD, 5% w/v PEG4K	diffraction
MPD H5	0.1 M Imidazole.HCl pH 8.0, 30 % (w/v) MPD	diffraction
MPD H6	30 %(w/v) MPD, 20 %(w/v) Ethanol	none
MPD H10	0.1 M HEPES sodium salt pH 7.5, 47 %(w/v) MPD	diffraction
MPD H11	47 %(w/v) MPD, 2 %(w/v) tert-Butanol	none
MPD H12	50 %(w/v) MPD	diffraction

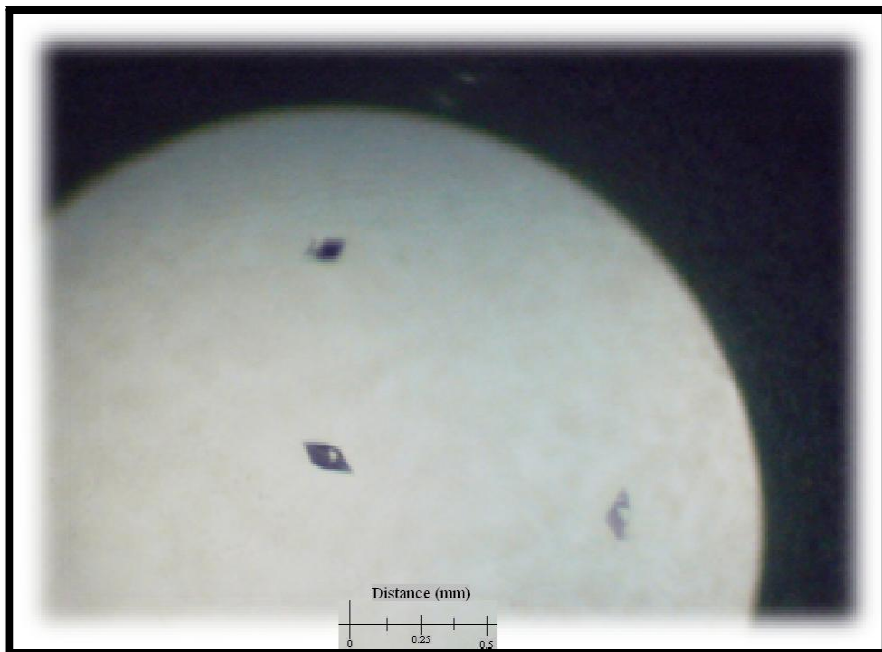


Figure 4.10 - Photograph of crystals within well 88 (H4) in the MPD screen. The largest of the three crystals (bottom left) measures 0.21 mm x 0.07 mm x 0.07 mm

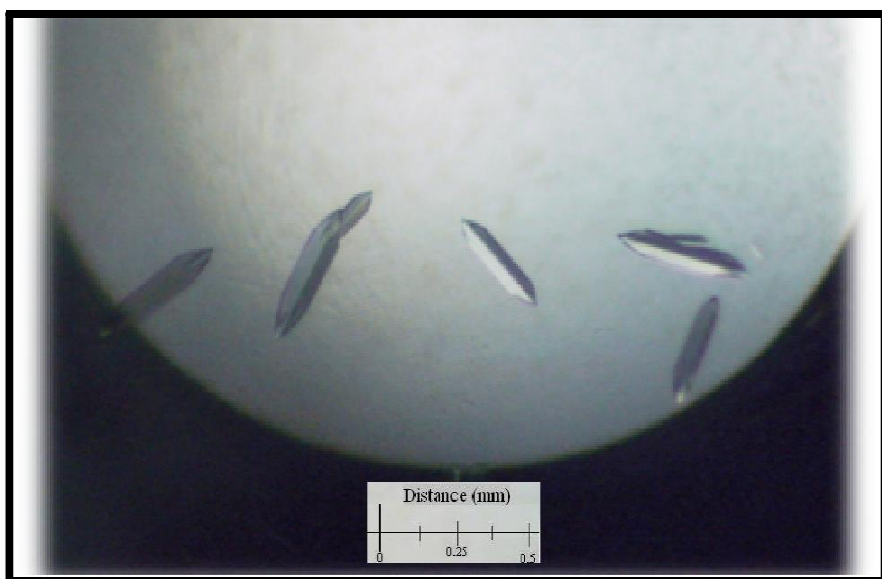


Figure 4.11 - Photograph of Optimized Crystallization condition (35% MPD, 7.5% PEG4000 and 0.1M HEPES buffer pH 7.5) in a sitting drop well. The crystal with a star next to it measures 0.34 mm x 0.068 mm x 0.068 mm.

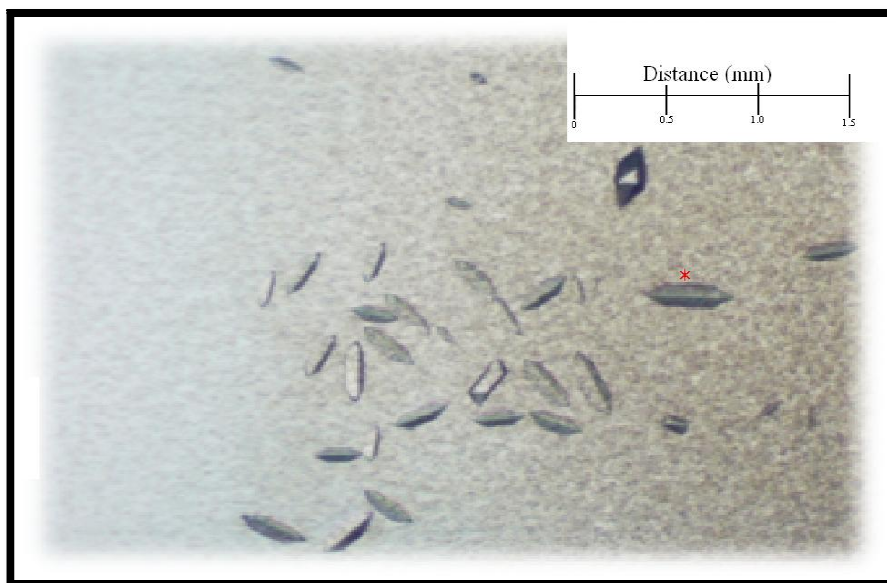


Figure 4.12 - Photograph of Hanging Drop Crystallization condition (35% MPD, 7.5% PEG4000 and 0.1M HEPES buffer pH 7.5) A protein film is noticeable in this picture because the photo was taken following exposure to open air to collect crystals for diffraction. Crystal with a red star beside it measures 0.51 mm x 0.11 mm x 0.11 mm.

The crystals produced in the optimized condition were diffracted at the APS in Chicago. The result of this diffraction was an incomplete data set with diffraction near 2.5 Å. The detected diffraction signal was not an improvement on previous native crystals which were diffracted. However, the key result is that the crystals grown utilizing the protein produced from the pMD80L construct diffracted at a resolution that was below 3.0 Å. This resolution would be sufficient for utilization with MAD phasing if the crystals contained SeMet rather than methionine.

4.2.7 Growth of pMD80L in Minimal Media

MenD protein from the pMD80L construct was then expressed in Nova Blue cells grown in minimal media containing methionine, not SeMet. A brief search of various protocols for SeMet expression yielded six separate protocols with slightly varied reagents for carbon and

nitrogen sources. Some of the preparations also involved the inclusion of the other nineteen amino acids and/or the four nucleotides as well. Of the six protocols, the successful protocol for expressing soluble MenD in minimal media was the condition that included both nucleotides and the amino acids in the mixture (Hendrickson 1990). Prior to induction with IPTG there is almost no expression of any protein, however upon induction the protein of interest is expressed but also appears to be the only protein being expressed within the soluble fraction (Figure 4.13).

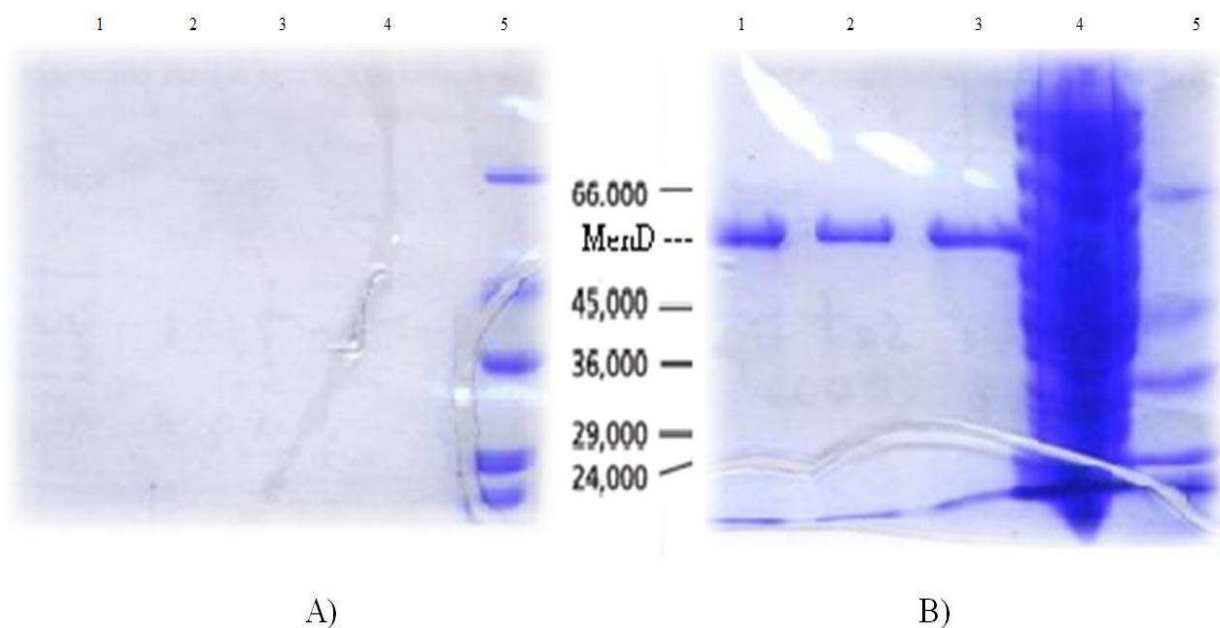


Figure 4.13 - Expression of MenD from pMD80L within minimal media A) SDS-PAGE gel showing no protein production prior to induction with IPTG. B) SDS-PAGE gel of three identical preparations following IPTG induction, harvesting, sonication and lysis of cells but not yet purified. Lane 4B contains cellular pellet showing insoluble protein and Lane 5A/5B are low molecular weight selectable marker.

4.2.8 Purification of MenD from pMD80L in Presence of TCEP

The protein sample acquired from the minimal media preparation was then purified using a nickel chelating column that is utilized to purify histidine tagged proteins. The procedure was intended to show that SeMet protein can be purified using a similar procedure to that of native

protein. The only difference between the two procedures is the presence of 0.2 mM TCEP in all of the buffers utilized to bind, wash and elute the column. TCEP is a reducing agent that stabilizes SeMet protein in solution and also does not reduce the nickel metal ions on the chelating column. ^(Bergendahl 2002) The results of the purification in the presence of TCEP (Figure 4.15) were very similar to that of the results seen in the absence of TCEP (Figure 4.14). From this result it is expected that SeMet protein will be able to be purified using the same procedure used with the native protein.

4.2.9 Expression of SeMet Protein from the pMD80L construct

A single SeMet preparation was done utilizing a high yield 12-L SeMet preparation kit, which had proven to be very successful for other proteins within the lab. The growth of cells within this medium was far slower than in any other expression preparation. In order for the cultures to grow and reach an OD₆₀₀ value above 0.3 a quantity of cells estimated to be more than ten times the quantity of cells used in any other preparation was required. The time period for this level of growth to occur exceeded twenty-four hours, however, once the log phase of growth was entered the bacterial cells behaved similarly to that of cells within LB media. Induction with IPTG took place at OD₆₀₀ equal to 0.4, as was done previously. The cells were grown for an additional eight hours, rather than the standard four hours. Despite the extra time given, a smaller quantity of cells was collected from the 1.0 L culture than was often collected in a 100 mL culture in LB media. The His-tag purification procedure also shows a smaller peak in the elution stage than in other analyses, but was nonetheless still present (Figure 4.16).

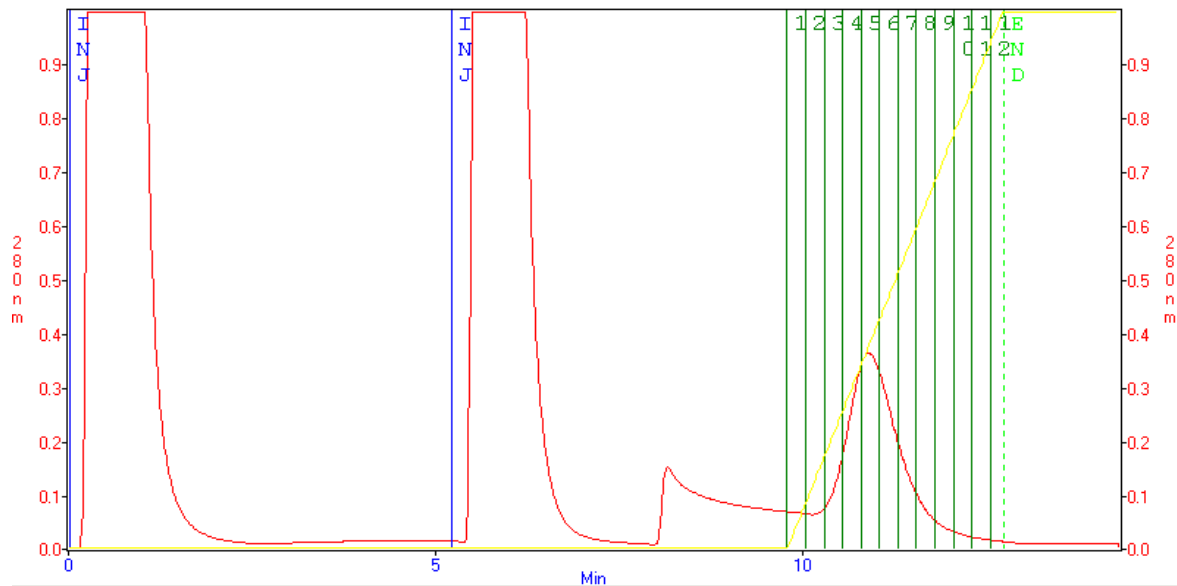


Figure 4.14 - Output chromatogram of purification of protein from pMD80L plasmid; Note the magnitude of the peak A_{280} in the stepwise gradient elution reaches nearly 0.4. Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.

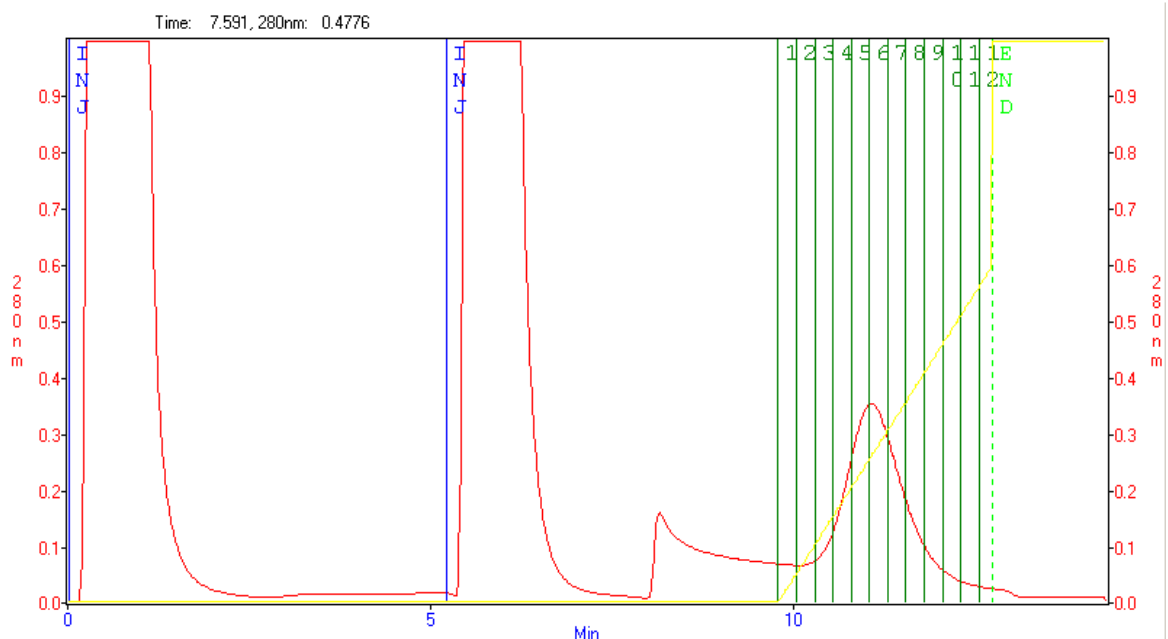


Figure 4.15 - Output chromatogram of purification of protein from pMD80L plasmid in the presence of 2 mM TCEP; protein sample used was identical to the sample used to produce Figure 4.11. Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.

Upon analysis with UV/Vis spectroscopy at 280 nm it was found that there was 9.6 mg of protein produced from the expression using the kit. However, upon undergoing dialysis the protein precipitated. The exact reason that this occurred is not understood but it is possible that the concentration of TCEP added to the dialysis buffer was inadequate. The sample was then filtered and concentrated and protein concentration measured again. It was found that less than 1.0 mg of protein was found in the remaining 2 mL of sample.

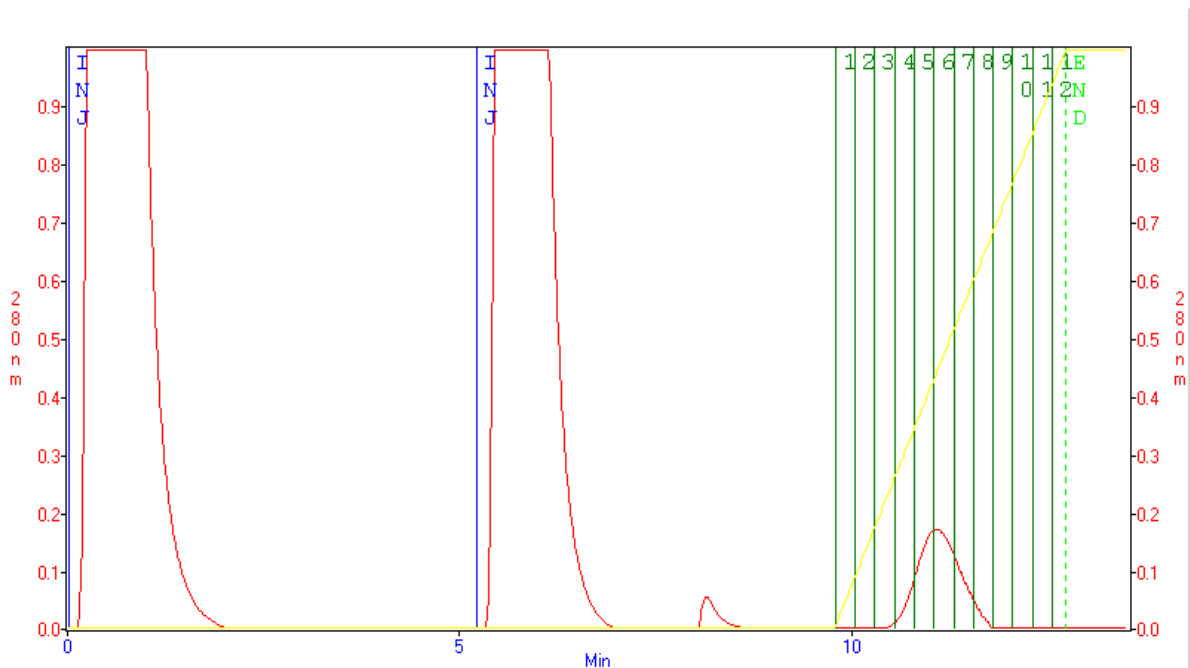


Figure 4.16 - Output chromatogram of purification of SeMet protein from pMD80L plasmid in the presence of 2 mM TCEP Protein was injected onto the column at 0 and 5 minutes. The wash buffer began at 8 minutes and the elution buffer gradient began just before 10 minutes.


4.2.10 Crystal Trials of SeMet pMD80L Protein

With the limited volume of SeMet protein remaining and the low concentration found (1.0 mg/mL), crystal trials were attempted with what was produced from the expression procedure. The concentration of protein was approximately 1.0 mg/mL, about one tenth of what was used when crystals were produced in native protein. The condition used was the same as was used for optimized native protein crystal trials (35% (w/v) MPD, 7.5% PEG4K and 0.1 M HEPES pH 7.5) with the addition of 4 mM TCEP to the well solution, such that the final concentration of TCEP in the drop would be maintained at 2 mM. The sitting drop technique was used with 1 μ L of protein to 1 μ L of precipitant within the drop. The result of this was no SeMet crystals; likely due to the low protein concentration.

4.3 Structure-Based Sequence Alignment of ThDP-Dependent Enzymes

4.3.1 Structure-based sequence alignment of ThDP-Dependent Enzymes using BAL (2AG0) as the Search Model

A structure based sequence alignment was completed on the enzymes within the DC family using BAL as the search model in an attempt to enhance our understanding of the similarities and differences between these enzymes. The online program FUGUE (Shi 2001) was used to perform the alignment and the format displayed is an output from the online program ESPript (Gouet 1999). The result is seen below in Figure 4.17. The focus of the discussion will be on variations that exist between the enzymes within the family in respect to key residues and sequences. Among these include the glutamate residues in the Pyr domain responsible for the activation of ThDP and the GDG-X₂₄-NN binding motif, which is responsible for Mg²⁺ binding and ThDP binding.

The glutamate residue that is thought to be responsible for ThDP to be activated, and form the ylide is said to be invariant (Frank 2007). Identified on the sequence alignment with a  it becomes noticeable that the glutamate is conserved across all ThDP enzymes. The results are in agreement with the research done previously with ThDP dependent enzymes.

The GDG-X₂₄-NN binding motif, which is responsible for Mg²⁺ binding and ThDP binding, is highly conserved within the ThDP dependent enzymes but is not invariant. The motif is conserved in benzaldehyde lyase (BAL, PDB code: 2AG0) (Mosbacher 2005), indolepyruvate decarboxylase (IPDC, PDB code: 1OVM) (Schutz 2003), benzoylformate decarboxylase (BFD, PDB code: 1BFD) (Hasson 1998), carboxyethyl arginine synthase (CEAS, PDB code: 1UPA) (Caines 2003), pyruvate decarboxylase from *Zymomonas mobilis* (ZPDC, PDB code: 1ZPD) (Dobritzsch 1998) and pyruvate decarboxylase from *Saccharomyces cerevisiae* (YPDC, PDB code: 1PVD) (Arjunan 1996).

For acetohydroxy acid synthase (AHAS, PDB code: 1JSC), the ThDP binding motif is slightly altered to GDA-X₂₄-NN. The mutation of the glycine to an alanine is a conservative mutation. Also, considering that the residue side chain is not involved in binding of ThDP or Mg²⁺, but rather the main chain oxygen, the mutation is rather insignificant.

The sequence of oxalyl-CoA decarboxylase (OCDC, PDB code: 2JIB) also indicates a discrepancy within the common GDG-X₂₄-NN motif that is present for binding of ThDP and the magnesium ion. Instead there is a GDS-X₂₄-NN present within the sequence of OCDC. However, once again this seems like a relatively minute alteration to the enzyme binding site of ThDP since it is the main chain carbonyl of that residue that contributes to binding and not the

side chain itself. The serine residue is larger spatially than glycine which is commonly seen or alanine as is present within AHAS. However, analysis of the residues present within AHAS, BAL and OCDC indicates the side chains and main chain oxygen are oriented similarly.

The sequence of pyruvate oxidase (POX, PDB code: 1POX) has a slight alteration to the GDG-X₂₄-NN binding motif to GDG-X₂₄-TN. The exchange of the asparagine residue to a threonine is again a conservative alteration to the motif since only the second asparagine residue is involved in Mg²⁺ binding within POX. The motif was referred to as GDG-X₂₅-N within the POX research article (Muller 1993). This was the convention at the time for displaying the motif since the first three models within the family of ThDP enzymes (POX, ScPDC and the model of transketolase) do not share alignment of the second asparagine residue. Muller's article also states that the glutamine (Gln476) adjacent to Asn474 within the motif, contributes to the Mg²⁺ binding. This contribution is filled by the second asparagine residue within the rest of the ThDP dependent enzyme family.

As stated within the introduction section acetolactate synthase (ALS) does not contain the asparagine portion of the GDG-(X₂₄)-NN motif that is conserved for binding of ThDP and Mg²⁺. This is clearly observed in the structure based sequence alignment of the ThDP dependent enzymes (Figure 4.17). It is notable that this is the only enzyme within the family that does not contain GDG-(X₂₅)-N motif. The discussion of this alteration (Pang 2004) mentions that the aspartate side chain oxygen (Asp474) and the backbone oxygen of Gly476 contribute to the binding of the diphosphate within the PP domain in the stead of the common NN motif.

```
BAL_mono 1 10 20 30 40 50
BAL_mono A.....MIT...GGELVVRTLIKAGVEHLFGTHGARIHDTFQACLDHVFPIIDTRHF...AAAGHAAEG
10ZG:A|PDB .....
```

```
BAL_mono 60 70 80 90 100 110 120
BAL_mono YARAGAKLGVAVLVTAGGFTAVTPANANWDRTPVILFVLTGSGALRDEETN...TLQAG...IDQVAMA
10ZG:A|PDB VGRITGKAGVALVTSQPCCSNLITGMATANSRGDPVVALLGGAVKRAKRAK...QVHQ...MDTVAMF
```

```
BAL_mono 130 140 150 160 170 180
BAL_mono RPITKWHRYVMAATEHPRLVMOA...RAALSAPGRPVLDL...WDILMNQIDEDSVIIPDLVLS
10ZG:A|PDB SPVTKYAIIEVTAPOALAEVVSNAFRAAEQGRPGSAFVSLF...QDQVQDGPVSGK...VLPASGAP
```

```
BAL_mono 190 200 210 220 230 240
BAL_mono AHGARP...P...D...Q...A...L...R...K...R...P...V...V...G...S...E...A...S...T...A...R...K...T...A...L...S...A...P...A...T...G...V...P...F...A...D...Y...E...G...L...S...M...L...S
10ZG:A|PDB QMGARP...DA...ID...Q...V...K...I...L...A...Q...A...K...N...P...I...F...L...L...G...L...M...A...S...P...E...N...S...K...A...L...R...R...L...L...E...T...S...H...I...P...V...T...S...T...Y...Q...A...A...G...A...V...N
```

```
BAL_mono 250 260 270 280 290 300
BAL_mono GLFDAMRGGGLVQNLVYFAKADRA...P...D...L...V...M...L...G...A...R...F...G...L...N...T...G...H...S...G...Q...L...I...P...H...S...A...Q...V...T...Q...V
10ZG:A|PDB QDNFSRFAGRVLGNFNNQAGDRLLQLA...D...L...V...I...C...I...G...Y...S...P...V...E...Y...E...P...A...M...W...N...S...G...N...A...T...L...V...H...I
```

```
BAL_mono 310 320 330 340 350 360
BAL_mono DFDACEGRVTOGIAIGIVADVGGTIRLAQAQATAQDAAWPDRGDWCAKVTDLAQER.YASIAA.KSSSE.H
10ZG:A|PDB DVLPAVIERNYTPQVELVGDVAGTLLNLAQNIIDHRLVLSPPAAEILRDRQVREL.LDRRGA.QLN.Q.F
```

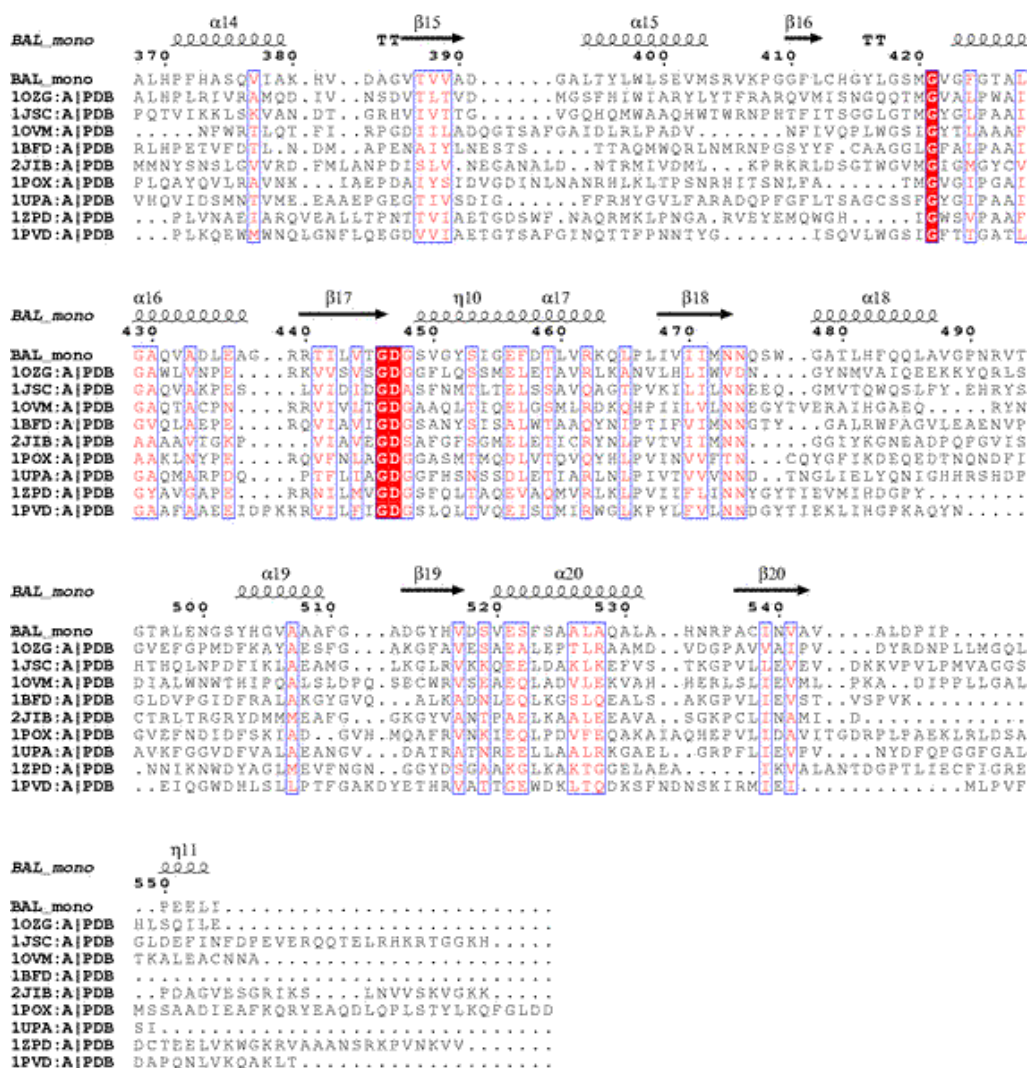


Figure 4.17 - Structure based sequence alignment of the DC family with benzaldehyde lyase as the structural search model

4.3.2 Structure-Based Sequence Alignment with MenD Modeller Model

A structure-based sequence alignment can be used to evaluate the validity of a theoretical model since it takes into account both the secondary structure features as well as the primary sequence features. If the MenD Modeller model were to show no discrepancy in the alignment of the key elements of the ThDP dependent enzymes binding motifs and conserved residues then

perhaps some determination could be made in regards to the shape of the active site, the determination of catalytically contributing residues and the binding of ThDP and Mg^{2+} .

The structure based sequence alignment of the *Modeller* model produced mixed results (Figure 4.18). When compared to the result of Figure 4.17, it is notable immediately that the invariant glutamate no longer is aligned amongst the entire list of enzymes. This could most likely be improved with a small manual alteration to the alignment of POX, since the glutamate required for invariant alignment is only one residue to the right. It was thought it was possible that the glutamate in POX was not the first residue in the alpha helix, as is the case for the rest of the enzymes, but analysis (Figure 4.19) indicates that this is not the case.

To confirm that the conserved glutamate was in the proper orientation within the model, ThDP from benzaldehyde lyase was superimposed into the ThDP binding site. The superimposition of ThDP shows that Glu55 within MenD is in fact in the orientation expected (Figure 4.20). Furthermore, there appears to be an arrangement of charged residues between the pair of active sites that may be indicative of a 20 Å tunnel that allows for catalytic duality (as seen in Figure 1.11) and non-equivalent active sites (Figure 4.21).

$\beta 1$ $\alpha 1$ $\beta 2$ $\eta 1$ $\alpha 2$ $\beta 3$

```

mend 1 10 20 30 40 50 60
mend SVSAFNRWA.....AVILEALTRHGVRHICIAFGSRSTPTLAAAEENSFAFHHTHFDER
2AGO:A|PDB.....MAMITG.....G.....EHLFGIHGAIIDTIFQACLDHDPVIDTRH.EA
1OZG:A|PDB.....MDKQYVVRQWA.....HGRQVFGIAGKIDKVFDSLDSIRIIPVRH.EA
1JSC:A|PDB AFSFNVDPLEQPAEFSKLAKKLRAPDMDSFVGLTGTDTVFGYFGGAILPVDIAIHNSCKFNFLPKH.EA
1OVM:A|PDB.....MRTFYCVADY.....DHLFVGFQDYMLQFLDHVIDSPDICWVGCAN.EA
1BFD:A|PDB.....MASVH.....DTVFGNFGSNELPFLKQFPEDFRYTLALQ.EA
2JIB:A|PDB.....MSNDNDNVELTDGDTMYGVVGFITNLARMWQDDGQRFYSFRH.EA
1POX:A|PDB.....TNILADHLYGIFGGSTNSIMDALSABERDRHYIQVRRHE
1UPA:A|PDB MSRVSTAPSGKPTAAHALLSR.....GKVFVVVGRFASLDFDEVEGIDFVLTREHEFA
1ZPD:A|PDB.....MSYTVGT.....KHHFVAVAGDYMLVLDNLLLNKNMEQVYCCNEL
1PVD:A|PDB.....SEITLNTVFGIFGDFNLSLQDKIYEVEGMRWAGNANET

```

$\alpha 3$ $\beta 4$ $\alpha 4$ $\alpha 5$ $\beta 5$ TT

```

mend 60 70 80 90 100 110 120
mend GLGHLALGLARVSKQPVRVIVVTSTAVANLYPALIEAGLITGKTLITADRPPELIDCGANQAIRQ...
2AGO:A|PDB AAGHAIEGYARAGAKLQVALVTAAGGFTNAVPIANAWLDRTPVLFITGSGALRDDDETNTLOAGID...
1OZG:A|PDB NAAPMAAAVGRITGKAGVALVTSGGPCSNLITGMATANSEGDVVVAIAGGAVKRADAKQVHQSMDT...
1JSC:A|PDB GAGHMAIEGYARASGKPGVVLVTSGGPAGTNVVFPMADAFADGIPMVVITGCVPTSAIGTDAFQEAQDV...
1OVM:A|PDB NASYAADGYARCKGFAALLTTFCVGELSAMNGIAGSYAHHVPLHIVGAPGTAAQDGELLHHTLG
1BFD:A|PDB HAGYASIAGYIEGKPGVCLTVSAPFLNGVTSLAHATTNCFPMILISGSSEREIVDLQOQDYEM...
2JIB:A|PDB CVVGIADGYAQASRKPAFINLHSAAGTGNAMGALSNAWNSHSELIIVTAGCQTRAMIGVEA...LLT
1POX:A|PDB EVGMFAAAADAKLTGKIGVCFGSAGPGGTHLMNGLYDAREDHVPLALIGCFGTTGMNMDTFOHMNE...
1UPA:A|PDB GVAADVLLARITGRPQACWATLGGMTNLSTGIATSVLDRSPVIALAAG...SESHDIFPNDTHQCLED
1ZPD:A|PDB NCGFSIEGYARANGAAAVVITVSGALSAFDAIGGATAENLPLVILISGAPNNNDHAAGVHLHNLG
1PVD:A|PDB NAAYAADGYARIKGMSCIITTFVGEALSALNGIAGSYAHHVPLHIVGVPSISSQAKQLLHHTLG

```

$\alpha 6$ $\beta 6$ $\alpha 7$ $\beta 7$ $\eta 1$ $\beta 8$ TT

```

mend 130 140 150 160 170 180
mend PGMFAS..HPTHISILP.RP..TQDIPARWVSTTQHLGLTLHAGGVHINCF...FAEPLYGEMDDTGT
2AGO:A|PDB QVAMAA..FITKWAHRV.MA..TEHIP.RLVVMQAIRAALSAPR.GPVLDDIP...WDILMNQIDEDSV
1OZG:A|PDB VAMPSP..VTKYAIEVT.AP..DALA...EVVSNAFRAAEQGRFSGAFVSLQDV...VDGPGVSKV...
1JSC:A|PDB VGISRS..C.TKNVMV.KS..VEELPLR..INEAFEIA.TSGRPGFVLDLQKDV..TAAALRNP IPTKT
1OVM:A|PDB DGEFRHFYHMSEPI TVA.QAVLTEQNACYEIDRVLTTLRERRPGYMLI..FADV..AKKAATPPVNA...
1BFD:A|PDB NVDANLPRPLVKWSY..EPASAAEVP.RAMSRATHMASMAPO.GPVYLSVP...YDDWDKDA
2JIB:A|PDB QDMNVARPHCKASFRIN.S...IKDIEPIARAVRTAVSG..RFGGVYVVDLPAKL..FGQTSVEAKKL
1POX:A|PDB NPIYADVAD..YNVTAV.NAATLPHV...IDEAIRRAYA..HQGVAVVQIIPVDL.PWQQTSAED...
1UPA:A|PDB VAIVAFMSKYAVELQRF.HEITD...LVDSAVNAAMTEPVGPFISLVDLGSSEGDITTVFNPP
1ZPD:A|PDB KTDYHYQLEMAKNITAAABEATVTEEPAKIDHVTKTALREKKPVYLEIACNIASMPCAAFGPASA...
1PVD:A|PDB NGDPTVVFHRMSANITSETTAMITDIATAPAEIDRCARTVYVYQRP..VYGLDANLVDLNVPAKLQTPID

```

$\alpha 8$ $\beta 9$ $\alpha 9$ $\alpha 10$ $\beta 10$

```

mend 190 200 210 220 230 240
mend LSWQQRQLGDWQDDKFWLREAPRLESEKQRDWFWRCKRGVVVAGRMSABEGKKVAL..WQTLGWPLIG
2AGO:A|PDB IIPDLVLVLSAHG...ARPDADLDCALALLRKAERPVIIVLGSSEASRTARKTALSATVAATGVPVFA
1OZG:A|PDB .....LPASGAPQMGAAFD.DAIDCVAKLIACAKNPIFLGLMASQPENSKALRRLLETSHPVTS
1JSC:A|PDB TLPNSALNQL..TSRAQDEFVMSINKAAADLNLAKKPVLYVYVAGIINHADGPRLL.KELSDRAQTPVTT
1OVM:A|PDB .....LTHKQAHADSACLKAFRDAENKILAMSKRTALLADFLVLRHGLKHALQKVVREVMMAHT
1BFD:A|PDB DPQSHHLEDRHVSSVRLND...QDLDLVLKALNSASNPAVVLGPDVDAANANADCVMLLETKAPVWV
2JIB:A|PDB L.....FKPIDPAPAQIPAEADAIARAADLTKNAKRPVMLGKGAAVYACQDDETRALVEETGIFLPL
1POX:A|PDB ..WYASANNYQT..PLLPE.PDQAVTILQTLLAERFLIYYGIGARKACKELE..QLKTLIKPLMS
1UPA:A|PDB ANTPAKPVGVVADG..WQKAA...DQAAALLAEAKHPVLVVGAAAIRSGAVFAIRALERLINIPVIT
1ZPD:A|PDB .....LFNDEASDEASLNAAVDETLKFIANRQKVAVIVGSKLRAA.GAEAAVKPTDALG...
1PVD:A|PDB MSL.....KPNDAESEK.EVIDGILALVKKAKNPFVLDACCSQHDVKAETKKLIDLTQFPFV

```

$\eta 2$ $\beta 11$ $\eta 3$ $\alpha 11$ $\beta 12$

```

mend 250 260 270 280 290
mend D....VLSQTGQPLPCAD.....LWLGNAKATSELOQAQIVVQDLSGSLTGRKRLQ
2AGO:A|PDB DYE..GLSMLSGLPDAMRG.....GLVQNLVYSAKADAPDLVVICMLGARF.GLNTGH
1OZG:A|PDB TYQAAGAVNQDNFSRFAGR.....VGLFNQAGDRLLCLADLVICIGYSPVEYEPAM
1JSC:A|PDB TLQGLGSDFDQ.EDFPKSLDM.....LGMHGCAITANLAVONADLIIAVGARFD.DRVTG
1OVM:A|PDB MLMGKGIEDERQAGFYGT.....SGSASTGAVKBAIEGADTVICVGT...
1BFD:A|PDB .....APSAFRCFPFTRHPCF.....RGLMPAGIAIISLLEGHQDVIIVIGAVVF..RYHQ
2JIB:A|PDB MGMAKGLLPDNHPSAAATRAFALAQCQDVCVLIGARLNWLMQHGKXGKTWGDDELKKYVVIDIQANEMD...
1POX:A|PDB TYPAKGI.....VADRYPAYL.....GSANKVACKPANEALQAQDVVIVVGNYP...FAE
1UPA:A|PDB TYIAKGVLPVGHENLYG.....AVTGYMDGILNFPALQTMFPVPLVILVIG...YD
1ZPD:A|PDB .....GAVATMAAAKSFPPENALYIGTSWGEVSVYVGVKTKKKA.DAVIALARVFNDSYTTG
1PVD:A|PDB TPMGKGSISEQHPRYGGV.....YVGTLSKPFVVKLAVESADLIIISVVALLSDPNTGS

```

$\beta 13$ $\beta 14$ $\alpha 12$

```

mend 300 310 320 330 340
mend WQASCEPEEYWIIVDDIEG.....RLDP AHRG..RRLIANIAD.WLELHP.....AEKRQWPWCVE
2AGO:A|PDB GSGQLIPHSAQVQVDPD.....ACELGRLOGEA..LGIVAVVGG.TIEALAQATQDAANVDRGDCWAC
1OZG:A|PDB WNSGNAT...LVHIDVLPF...AYE.ERNYTPD..VELVGDLAG.TLNKLAQNIDRRLVLSPPQAAEILR
1JSC:A|PDB NISKAPAEARRAAEGRGGIHFVSPKNINKV..QTQIAVGGD.ATTNLGKMMSKIFPKERSSEWFA
1OVM:A|PDB .....RFTDTLTAGFTH...QLTPAQITIEV.QPHAAAVGDVWFTGIPMNQAIETLVLCKQHVHA
1BFD:A|PDB Y....DQFGQY...LKPGTALISVTCDF...L.EAARAPMGDAIVA...D.GAMASALANLVEE
2JIB:A|PDB ..SNQPIAAPVVGDIKSAV.....SILRKLKAGPKA.....DAEWTGA
1POX:A|PDB VSKAFKPNTRYFLQIDIDPA.....KLGKRAKHTDI.AVL.ADAQKTLAAIILAQVSEB...ESTPWNQAN
1UPA:A|PDB YAEOLRNSPMW..QKGIKKTIVRSFTVNFIPRVY.RFDVDDVTDVLAIVVHFETATASFGAKQRHIEP
1ZPD:A|PDB WT.....DIPDFKKLVAEPRSVVNGI.RFSPVHLKDYLTRL.....AQKVSKKTGS
1PVD:A|PDB FSYSYKTRN..IVEFHSDHMKIRNATPPGVQMKFVLOKLLINLADAAGKYK...VAVPARTPANAA

```

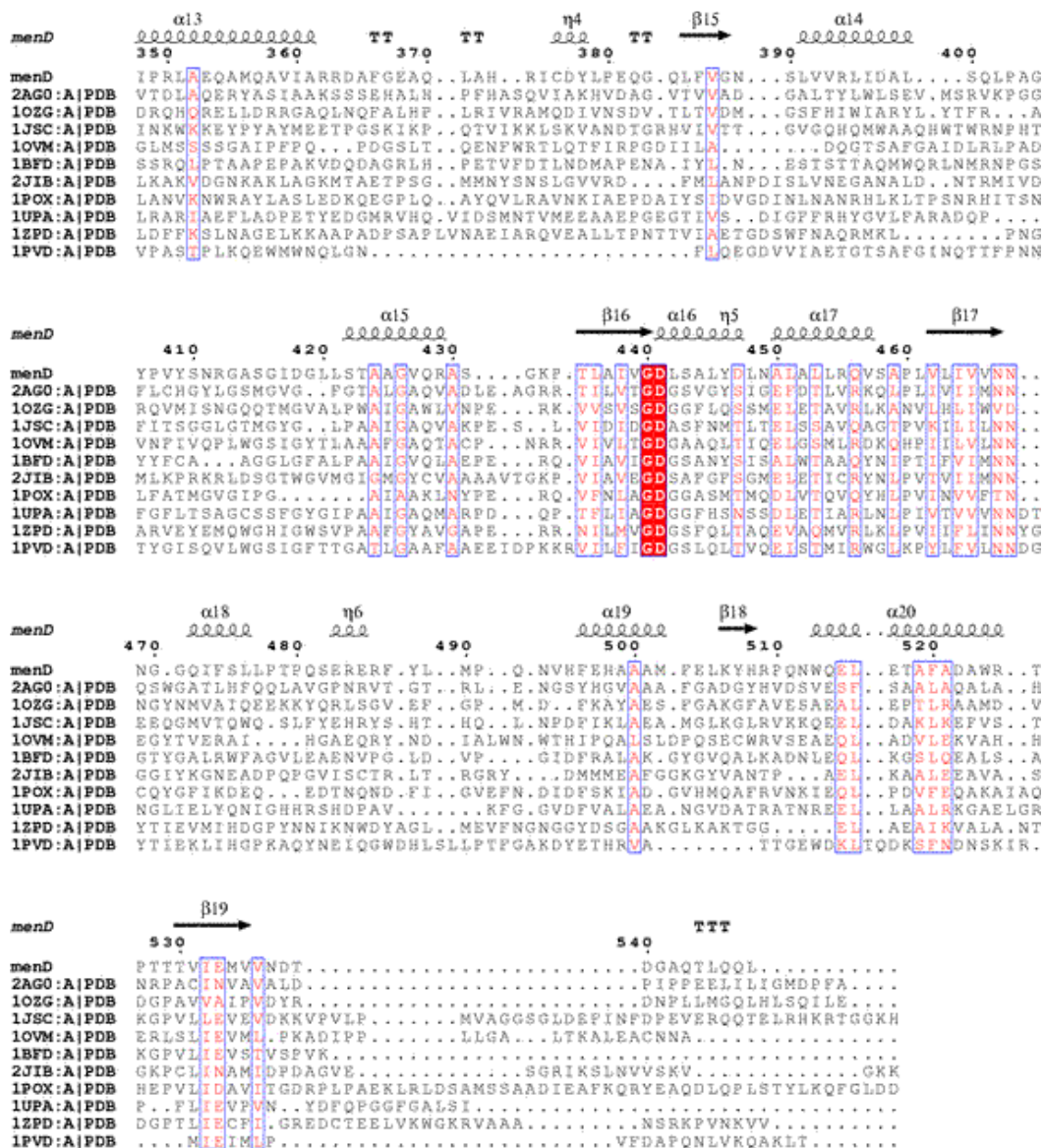


Figure 4.18 - Structure based sequence alignment of the DC family with the MenD Modeller model as the search model. It is important to note that the sequence number listed at the top of each segment indicates the amino acid number minus one. This is due to the lack of methionine at the beginning of the sequence. For example the invariant Glu55 of MenD would appear to be numbered 54 in this figure.

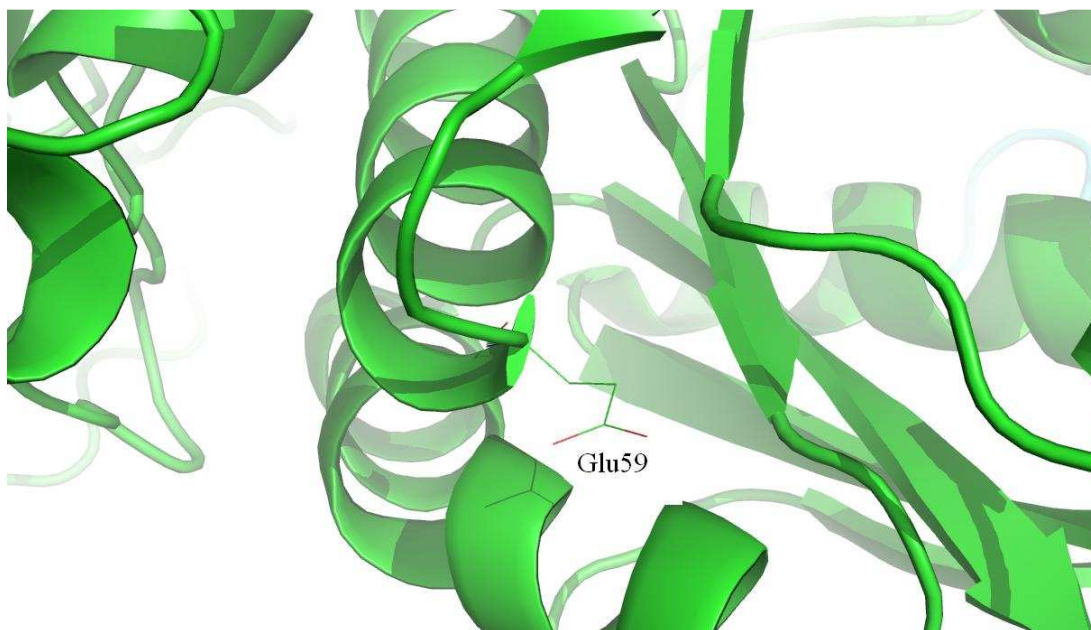


Figure 4.19 - Analysis of the IPOX model to ensure Glu59 is the first residue in the alpha helix as is the tendency for the invariant glutamate residue in ThDP-dependent enzymes.

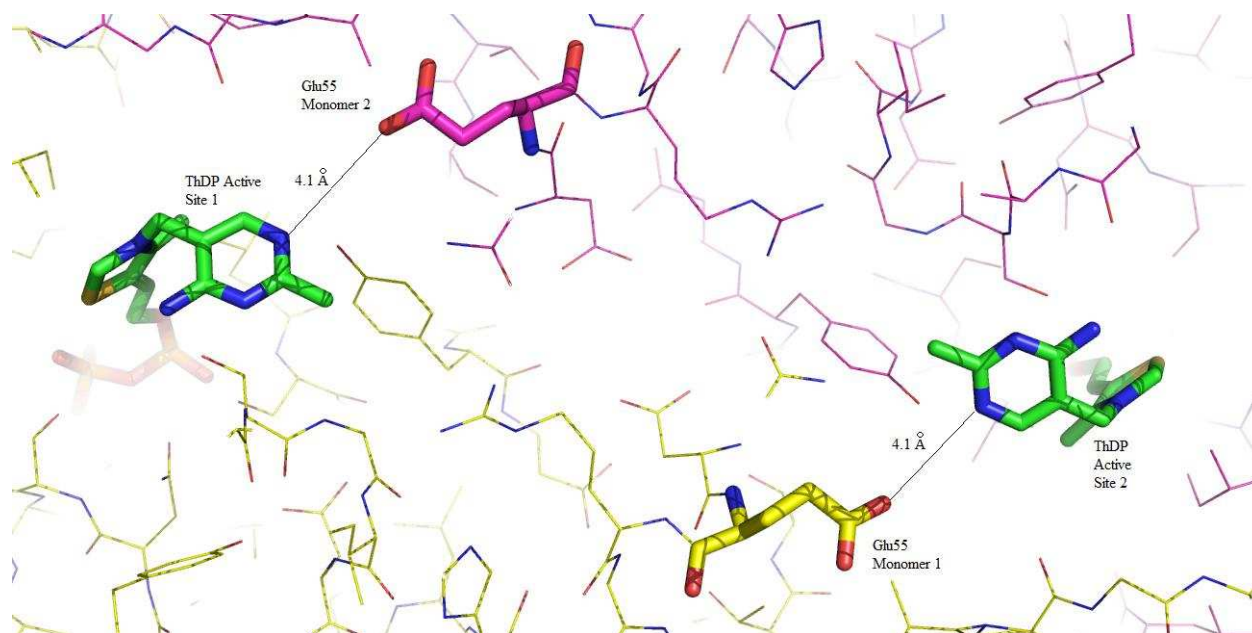


Figure 4.20 - Location of Glu55 within the MenD *Modeller* Model. The distance from the carboxyl group of glutamate to the N1' atom of the pyrimidine is 4.1 Å (measured with Coot) which is slightly out of range for hydrogen bonding. The orientation is as was expected though.

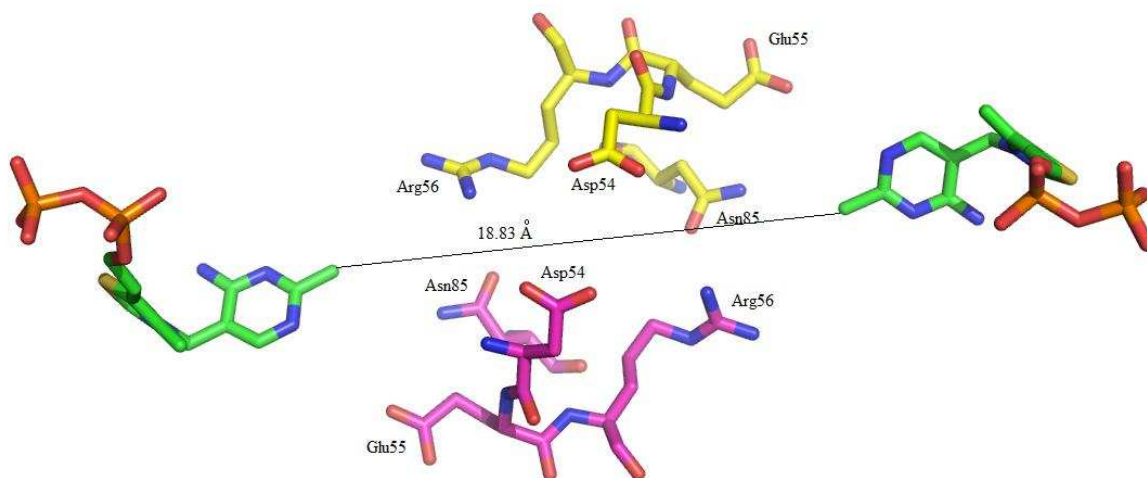


Figure 4.21 - The 18.83 Å tunnel between the two active sites within one of the dimer pairs of the MenD *Modeller* model. The charged residues between the pair of active sites may form a hydrogen bonding bridge that explains the non-equivalence of active sites often seen in ThDP-dependent enzymes (Figure 1.11).

The alignment of the GDG-(X₂₄)-NN motif using the MenD *Modeller* model remains the same as it was for the alignment using 2AG0. However, alignment of the MenD structure indicates that the GDG portion is not conserved similar to that of OCDC and AHAS. The implication is similar as it would be assumed that the backbone oxygen atom would contribute to binding and that the GDL sequence seen in MenD would have no implication on binding. The assumption is made that the large leucine residue will not point into the ThDP binding site but rather into the interior of the protein as was the case for the serine in OCDC (2JIB). Upon analysis (Figure 4.22) it is shown that Leu443 runs along the edge of the ThDP binding site and may contribute to induction of the V-conformation of ThDP upon binding. In most ThDP dependent enzymes, the main chain oxygen contributes to diphosphate binding; however this does not seem to be possible in the *Modeller* model since the main chain oxygen points away from the diphosphate. It is indeed possible that this unique sequence difference in MenD when

compared to other ThDP-dependent enzymes has an alternative role for ThDP binding and activation.

As was stated in the introduction, previous research done in our lab has implicated Ile418 as the residue responsible for the induction of the V-conformation of ThDP. Identification of this residue within the model was also done (Figure 4.23) and it shows that within the model this residue has no role in the binding of ThDP. This result is indicative that either the *Modeller* model is incorrect or that Ile418 has a very important internal structural role for MenD since mutagenesis of this residue results in significant loss in enzyme activity.

There are various other residues that have been implicated via mutagenesis studies to be contributing residues of catalysis of the MenD reaction. Among these is Asp442, which was implicated for Mg^{2+} binding. The MenD *Modeller* model shows this residue to be in direct proximity of the ThDP moiety (Figure 4.24). Also, Figure 4.24 and the structure based sequence alignment indicate that Asp442 is part of the GDL-X₂₄-N binding motif.

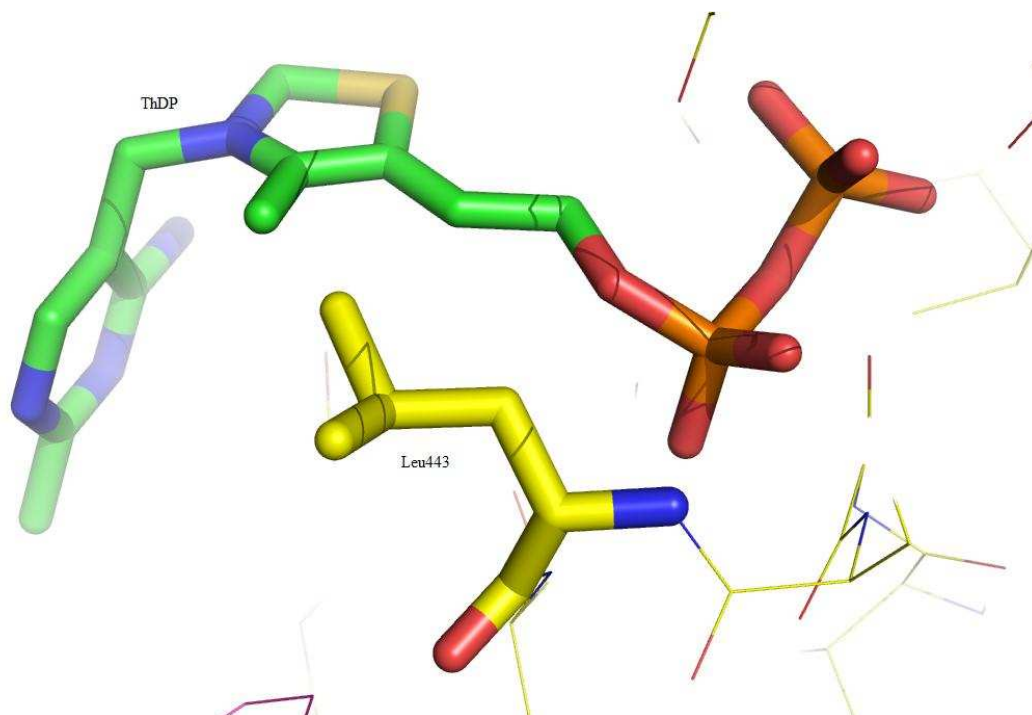


Figure 4.22 - Location of Leu443 within the MenD *Modeller* model; Note the orientation that runs along the edge of the ThDP binding pocket.

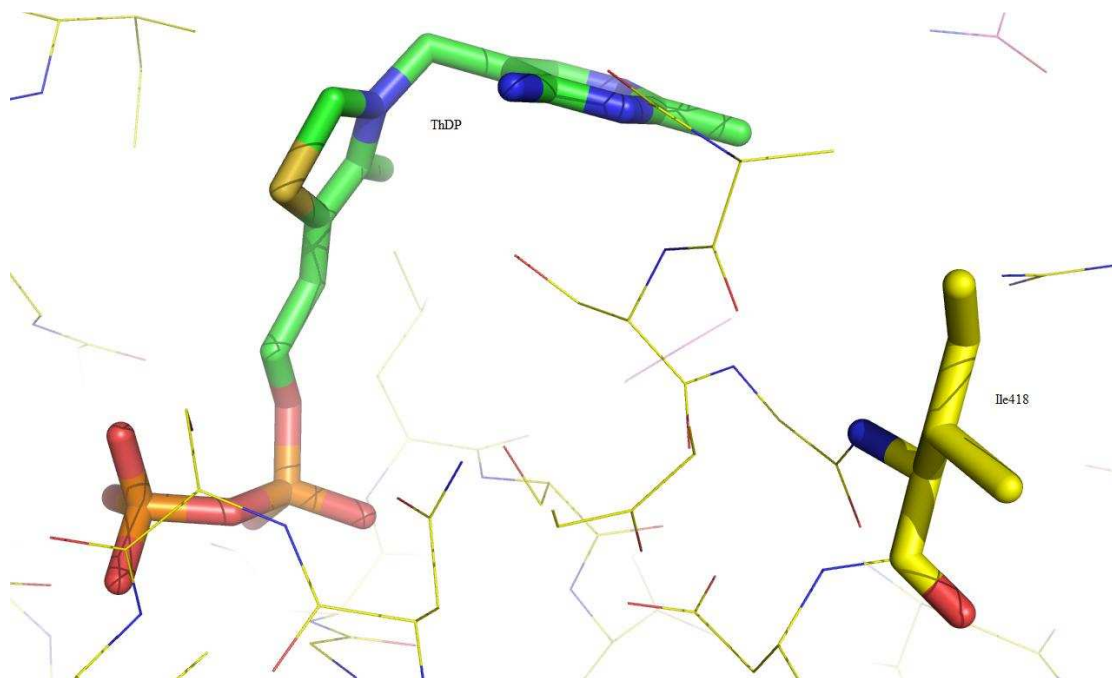


Figure 4.23 - Location of Ile418 within the MenD *Modeller* Model; Note the location is not in close proximity to the cofactor ThDP.

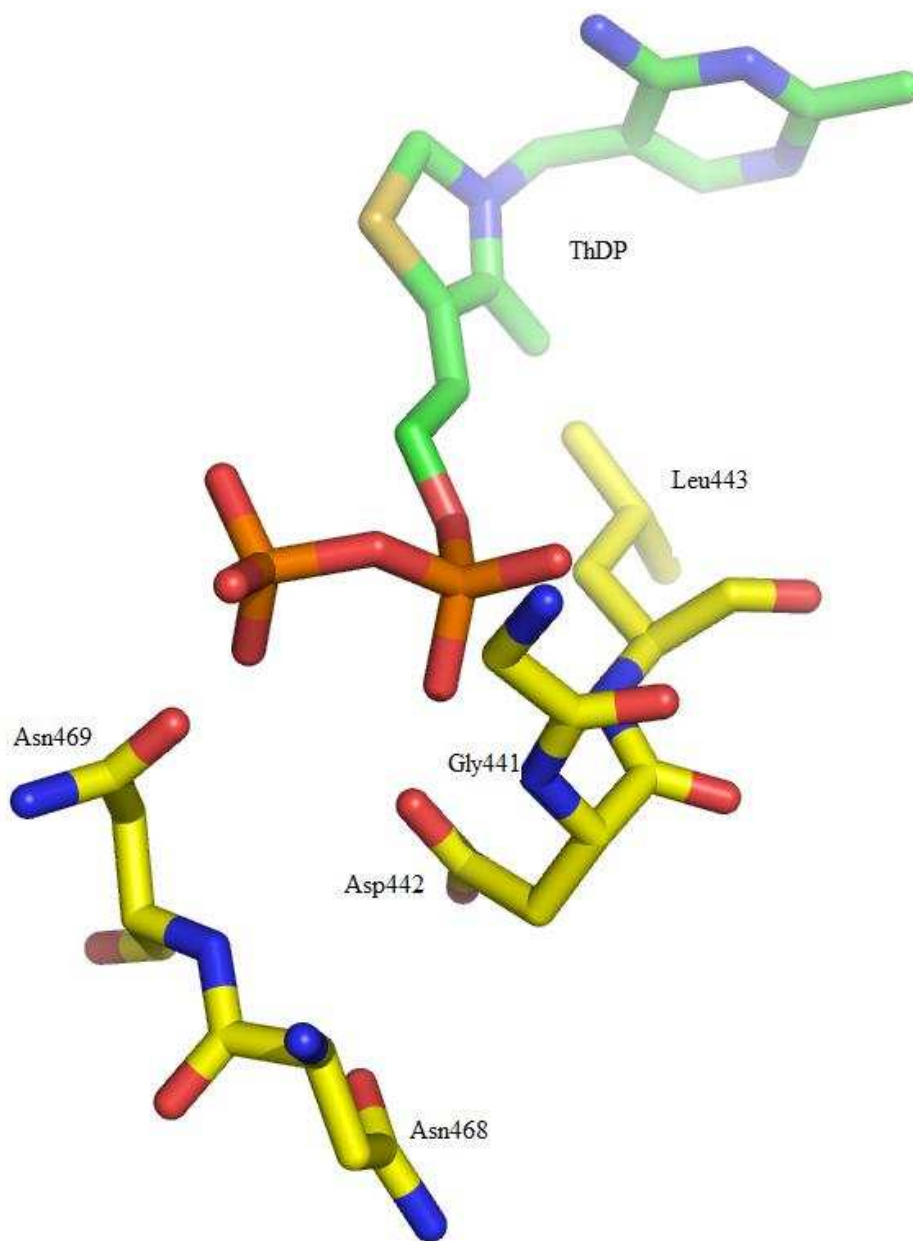


Figure 4.24 - Location of Asp442 within the MenD *Modeller* Model; Note the orientation that runs directly adjacent to the diphosphate and can be assumed to be in close proximity to the Mg^{2+} cofactor. Also note the presence of the remainder of the GDL- X_{24} -NN binding motif within this figure and how all of these residues are in direct proximity to the diphosphate of ThDP

4.3.3 MenD *Modeller* Model Active Site Residue Analysis

In addition to the residues that have been discussed thus far, a variety of other residues within MenD have been investigated with mutagenesis studies including Ser32, Arg33, Arg395, Gln118 which were all thought to participate in catalysis. Upon analysis of the MenD *Modeller* model it was found that these results were in agreement with the model with the exception of Arg395. Ser32, Arg33 and Gln118 point directly into the active site and are directly adjacent to the C2 position of the thiazolium ring (Figure 4.25) Although Arg395 is adjacent to the active site it is too far from C2 of the thiazolium ring to make the conclusion that it participates in catalysis. Along with Ser32, Arg33 and Gln118, the model indicates that the active site is also formed by Leu478 and Asn117 (also in Figure 4.25). These results are indicative of the validity of previous mutagenesis studies done within the Palmer laboratory (Macova 2005, Ho 2007).

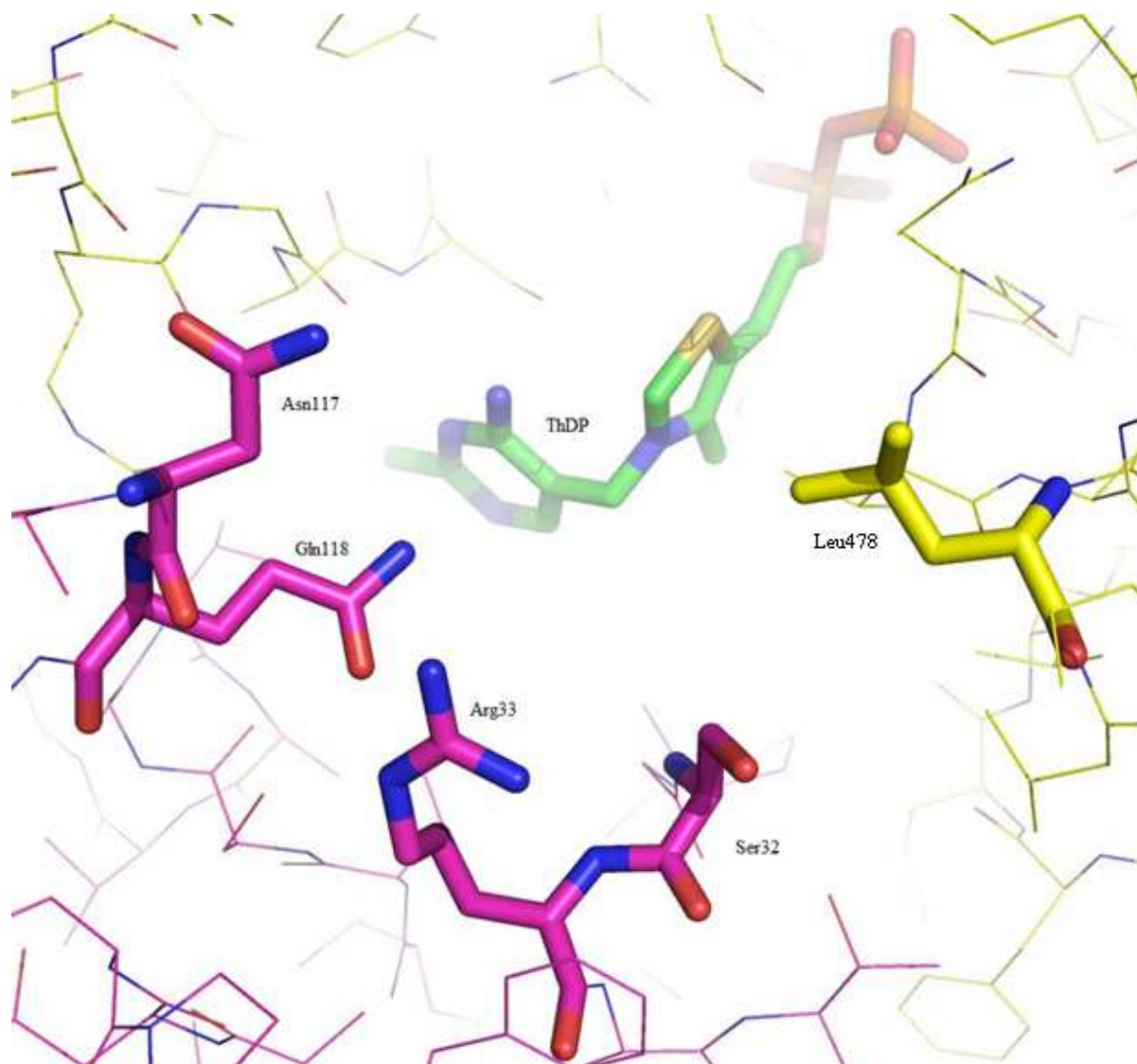


Figure 4.25 - The MenD active site residues according to the MenD *Modeller* Model; Residues thought to be capable of participating in catalysis/binding of substrate include (but are not limited to) Asn117, Gln118, Ser32, Arg33 and Leu478.

Chapter 5:

Concluding Statements and Summary

The results described in this thesis have provided the research project with a variety of contributions. The data set collected in 2005 was processed and it was determined that the holoenzyme data set was more suitable for molecular replacement strategies. It was shown that with the current models listed within the protein data bank, the use of molecular replacement is not a reasonable means to solve the phase problem for MenD with this data. A variety of software packages including MRBUMP, AMoRE and PHASER were utilized to achieve this task but none were successful. The final attempt to solve the phase problem via molecular replacement was done by generating a model of MenD with the program *Modeller*. Although this strategy was also unsuccessful in solving the phase problem for the holoenzyme data set, the model developed provides the opportunity to make a variety of implications in regards to the structure function relationship of MenD and the binding of ThDP.

With the results from the computational efforts with the data set provided it was necessary to develop a new system for producing SeMet MenD protein. This system was developed by cloning the MenD gene into a new plasmid (pQE80L), which was ideal to produce SeMet protein to be utilized for MAD phasing. The resulting plasmid construct (named pMD80L) was able to express active MenD enzyme, in both LB broth and minimal media. It was also possible to utilize previous methods of purification with pMD14b in the presence of TCEP to purify protein produced from pMD80L. Native crystallization conditions for pMD80L protein were determined and the crystals diffracted to 2.5 Å at the Advanced Photon Source in Chicago. SeMet protein was expressed with a SeMet kit, however, inclusion of SeMet was not confirmed.

Future work done on this project should focus on the expression, purification and crystallization of the SeMet derivative of MenD from pQE80L. The major project goal to solve the phase problem and produce an accurate model of MenD that could be submitted to the protein data bank was not completed. However, many steps have been made to make completion of this goal a reality in the near future. Further optimization of the crystal conditions may be necessary to produce diffraction quality crystals for the SeMet derivative protein. Although, if a MAD data set of less than 3.0 Å is produced the phase problem should be able to be solved. With the 1.9 Å holoenzyme data set already processed, a high resolution structure of MenD is achievable in the near future for this lab group.

During the preparation of this thesis the structure of MenD was published by a group from The University of Dundee in Scotland in the Journal of Molecular Biology (Dawson 2008). The structure was solved via single wavelength anomalous dispersion (SAD) phasing at 2.8 Å resolution using a SeMet derivative. The structural model was then improved to 2.5 Å using molecular replacement with a native data set.

The structure shows the general ThDP-dependent enzyme architecture with the three domains, each consisting of six parallel β -sheets sandwiched between several α -helices in an α/β supersecondary structural motif (compared to *Modeller* model in Figure 5.1). MenD was confirmed to be part of the DC family of enzymes and was shown to be most structurally similar to benzoylformate decarboxylase (C_{α} -RMSD = 2.54 Å) and oxalyl-CoA decarboxylase (C_{α} -RMSD = 2.56 Å).

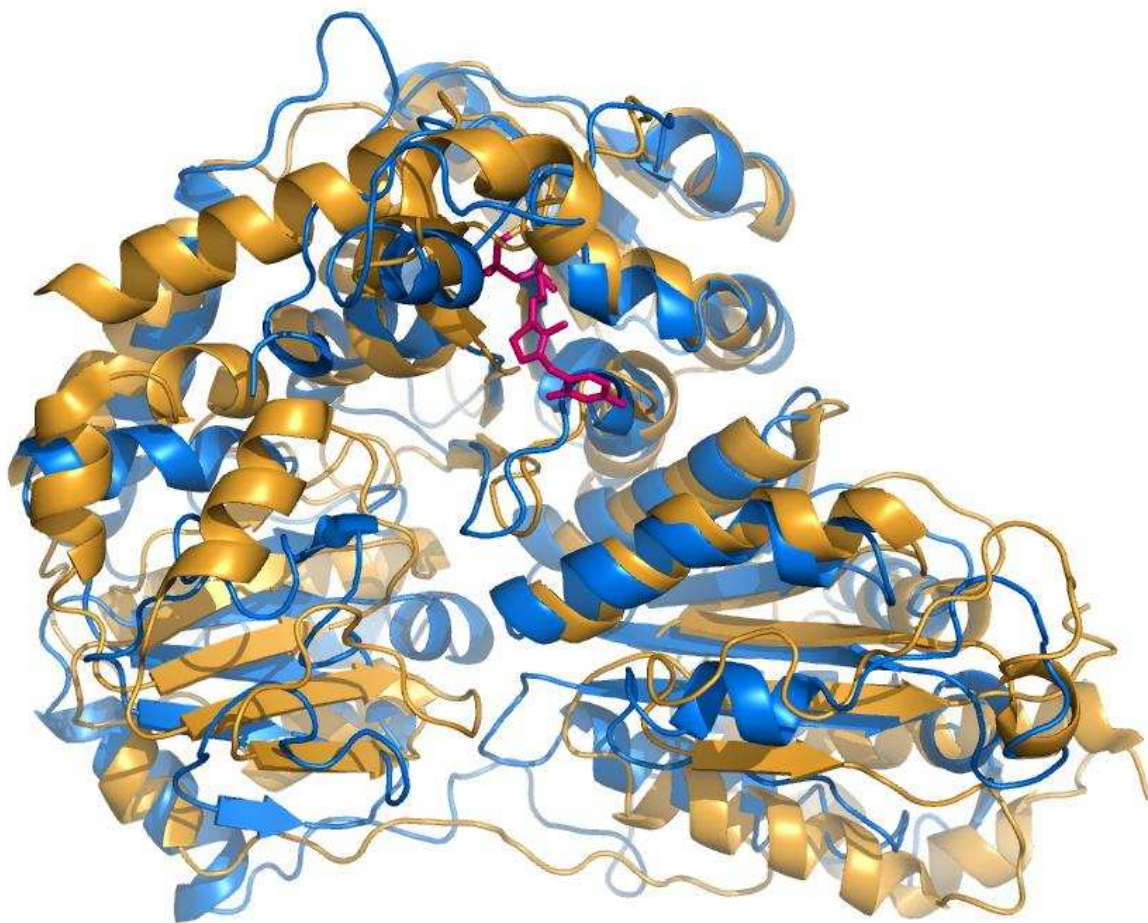


Figure 5.1 – The structural model of MenD (gold) superimposed over the *Modeller* model (blue). ThDP is shown as a stick model in pink. Noticeable differences are seen between the two models, specifically in the middle domain where the six parallel β -sheets do not overlap.

Several discrepancies are now seen in the *Modeller* model discussed in this thesis. The structural differences are viewed with the superimposition of the six parallel β -sheets within the middle domain and a variety of α -helices throughout the structure (Figure 5.1). Furthermore, the induction of the V-conformation is in fact caused by Ile418 (Figure 5.2) as was assumed from mutagenesis studies (Ho 2007) but could not be confirmed with the *Modeller* model (Figure 4.23).

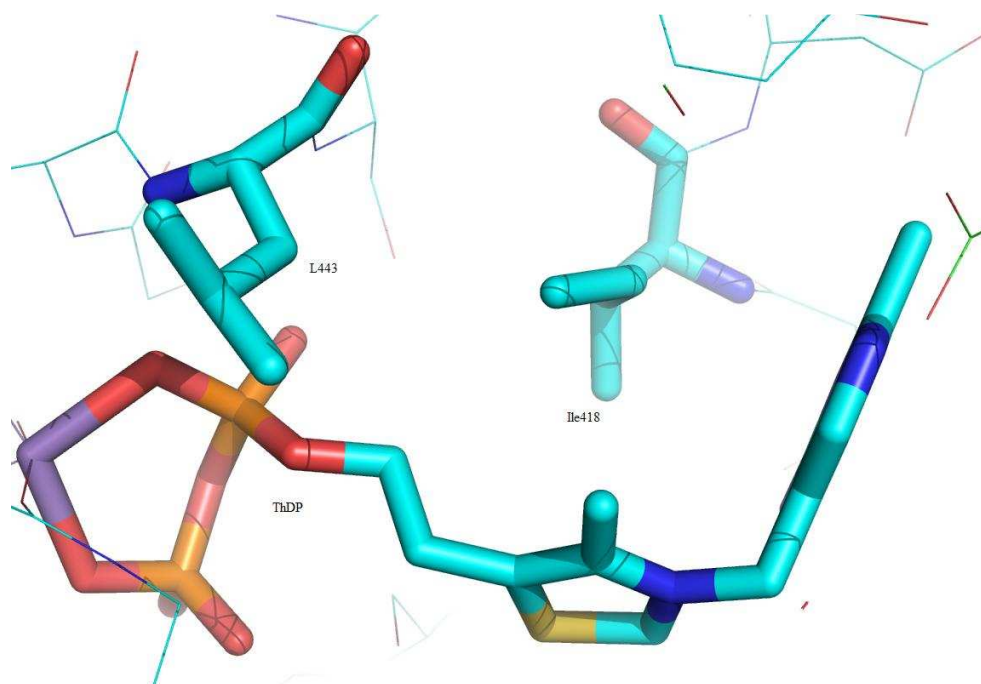


Figure 5.2 – Location of both Ile418 and Leu443 in the MenD model. This indicates that the V-conformation is induced by Ile418 and not Leu443.

Finally, significant differences are seen for the active site residues implicated to participate in substrate binding and catalysis (Figure 4.24). Upon analysis of the structure of MenD (Dawson 2008), it is notable that the catalytic residues implicated by the Palmer laboratory (namely Ser32, Asn117 and Gln118) are indeed present within the active site cleft (Figure 5.3) and could participate in catalysis. However, the majority of the active site is composed of basic arginine, lysine and asparagine residues that aid in the binding of MenD's acidic substrates, isochorismate and 2-ketoglutarate. Ile474, Phe475, and Leu478 are also present within the active site and are implicated in binding of the hydrophobic portion of isochorismate.

Despite several differences, there are certainly similarities seen when comparing the *Modeller* model to the structure of MenD. For instance, when comparing the locations and conformations of the residues involved in the ThDP-binding motif GDL-(X₂₄)-NN for the

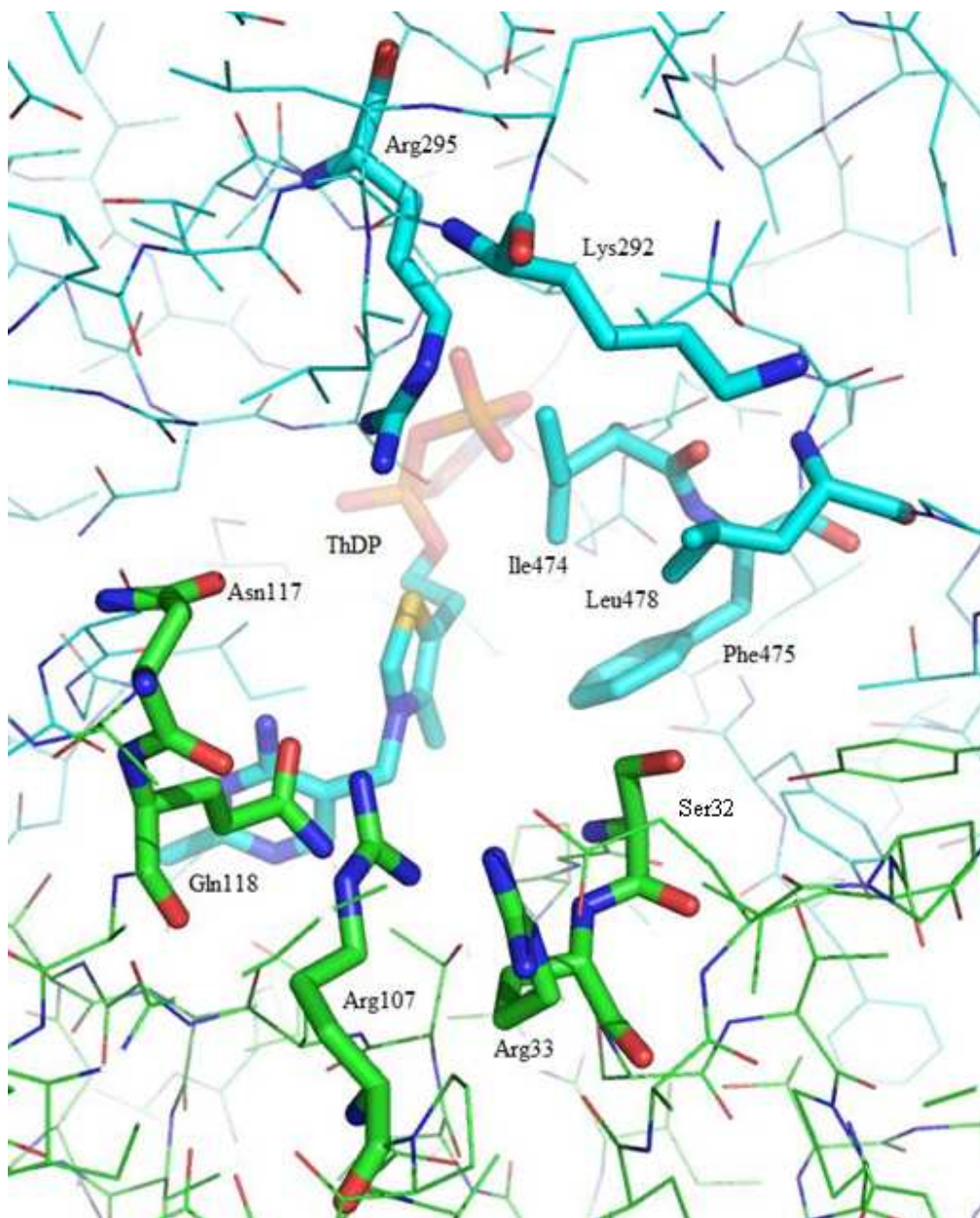


Figure 5.3 – The active site of MenD. Arg33, Arg107, Lys292 and Arg295 aid in the binding of MenD’s acidic substrates, isochorismate and 2-ketoglutarate. Hydrophobic contacts are made with the hydrophobic ring of isochorismate by Ile474, Phe475, and Leu478. Note that the active site is formed from a dimer, monomer A is shown in green and monomer B is shown in light blue.

structure of MenD (Figure 5.4) to that presented for *Modeller* model (Figure 4.24) they are almost identical. The structure of MenD also shows similar orientation of the residues implicated to form the 20 Å catalytic duality tunnel that shuttles a proton between the pair of active sites (Figure 5.5) to that seen in the *Modeller* model. However, the residues Asp54, Glu55, Arg56 and Tyr447 are thought to block access between the two active sites preventing the transfer of a proton (Dawson 2008).

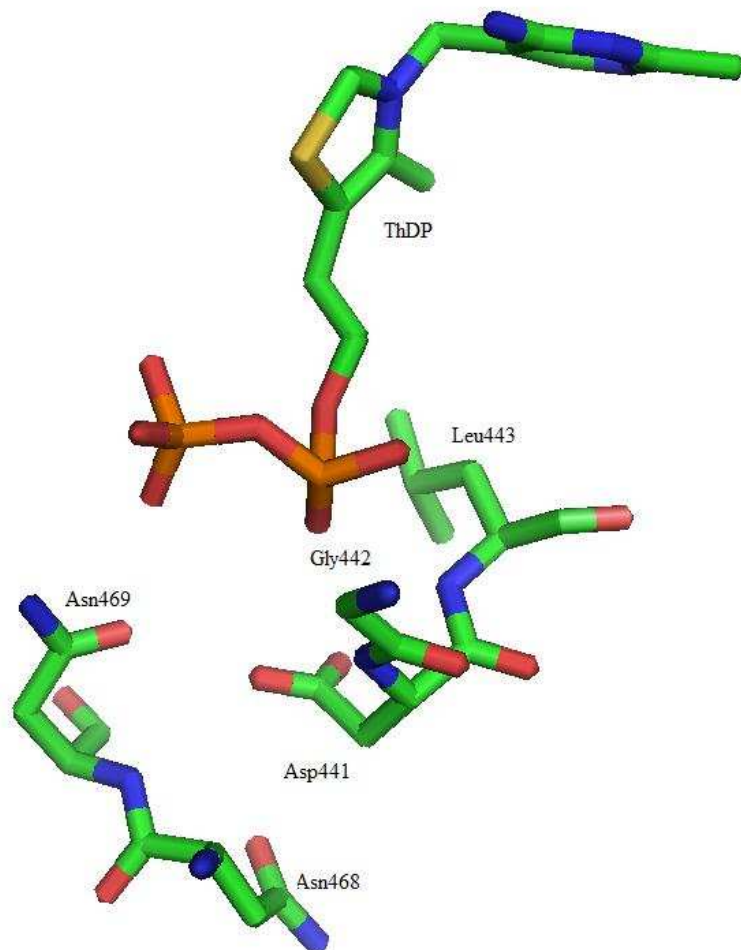


Figure 5.4 – ThDP-binding motif GDL-(X₂₄)-NN within the MenD structure.

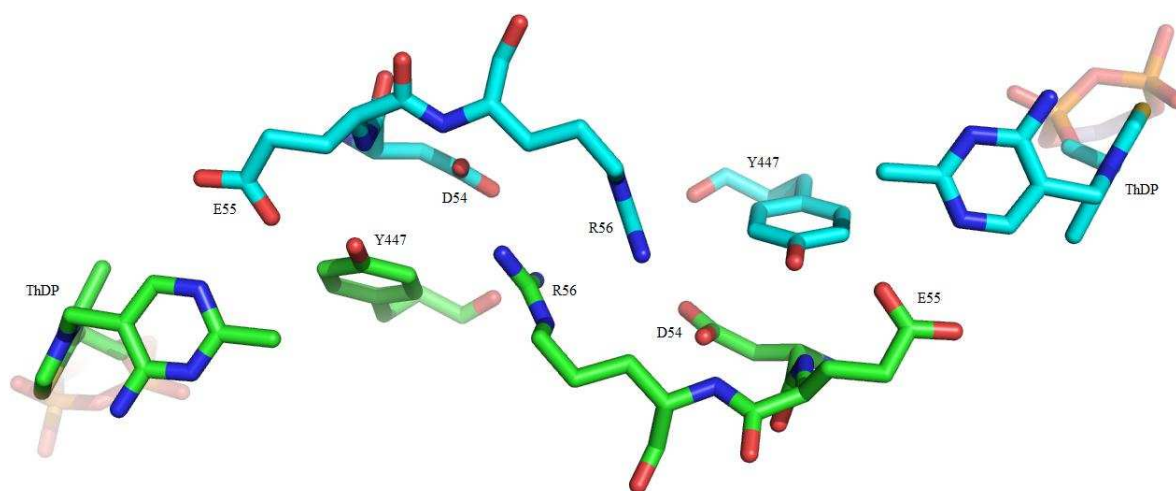


Figure 5.5 – The absence of a catalytic duality tunnel in the structural model of MenD. The residues Asp54, Glu55, Arg56 and Tyr447 block access between the two active sites.

Chapter 6: References

- Bentley, R., Meganathan, R. (1982). "Biosynthesis of Vitamin K (menaquinone) in Bacteria." Bacteriological Reviews **46**(3): 241-280.
- Bentley, R., Meganathan, R. (1983). "Vitamin K biosynthesis in bacteria - precursors, intermediates, enzymes, and genes." Journal of Natural Products **46**: 44-59.
- Bergendahl, V., Anthony, L.C., Heyduk, T., Burgess, R.R. (2002). "On-column tris(2-carboxyethyl)phosphine reduction and IC5-maleimide labeling during purification of a RpoC fragment on a nickel-nitrilotriacetic acid Column." Analytical Biochemistry **307**(2): 368-374
- Berthold, C.L., Moussatche, P., Richards, N.G.J. (2005). "Structural basis for activation of thiamine diphosphate-dependent enzyme oxalyl-CoA decarboxylase by adenosine diphosphate." Journal of Biological Chemistry. **280**(50): 41645-41654.
- Berthold, C. L., Toyota, C.G., Moussatche, P., Wood, M.D., Leeper, F., Richards, N.G.J., Lindqvist, Y. (2007). "Crystallographic Snapshots of Oxalyl-Coa Decarboxylase Give Insights Into Catalysis by Nonoxidative Thdp-Dependent Decarboxylases" Structure **15**(7): 853-861.
- Bhasin, M., Billinsky, J.L., Palmer, D.R.J. (2003). "Steady-state kinetics and molecular evolution of Escherichia coli MenD [(1R,6R)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase], an anomalous thiamin diphosphate-dependent decarboxylase-carboligase." Biochemistry **42**(46): 13496-13504.
- Bishop, D.H.L., Pandya, K.P., King, H.K. (1962). "Ubiquinone and Vitamin K in bacteria." Biochemical Journal **3**(83): 606-614.
- Bowen, T.L., Union, J., Tumbula, D.L. (1997). "Cloning and phylogenetic analysis of the genes encoding acetohydroxy acid synthase from the archaeon *Methanococcus aeolicus*." Gene. **188**(1): 77-84.
- Caines, M. E., Elkins, J.M., Hewitson, K.S., Schofield, C.J. (2004). "Crystal structure and mechanistic implications of N²-(2-carboxyethyl)arginine synthase, the first enzyme in the clavulanic acid biosynthesis pathway." Journal of Biological Chemistry **279**: 5685-5692.
- Candy, J.M., Duggleby, R.G. (1998). "Structure and properties of pyruvate decarboxylase and site-directed mutagenesis of the *Zymomonas mobilis* enzyme." Biochimica et Biophysica Acta – Protein Structure and Molecular Enzymology. **1385**(2): 323-338.

Cheng, T. (2003). "Activated protein C blocks p53-mediated apoptosis in ischemic human brain endothelium and is neuroprotective." Nature Medicine **9**: 338-342.

Daruwala, R., Kwon, O., Meganathan, R. and Hudspeth, M.E. (1996) "A new isochorismate synthase specifically involved in menaquinone (Vitamin K₂) biosynthesis encoded by the menF gene." FEMS Microbiology Letters. **140**: 159-163.

Dawson, A., Fyfe, P.K., Hunter, W.N. (2008). "Specificity and Reactivity in Menaquinone Biosynthesis: The Structure of *Escherichia coli* MenD (2-Succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexadiene-1-carboxylate synthase)." Journal of Molecular Biology: Article in Press

DeLano, W.L.(2002)The PyMol Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA

Demir, S., Pohl, M., Janzen, E., Muller, M. (2001). "Enantioselective synthesis of hydroxyl ketones through cleavage and formation of acyloin linkage. Enzymatic kinetic resolution *via* C-C bond cleavage. Journal of the Chemical Society, Perkins Transactions **1**: 633-635

Dobritzsch, D., Konig, S., Schneider, G., Lu, G. (1998). "High Resolution Crystal Structure of Pyruvate Decarboxylase from *Zymomonas mobilis*." The Journal of Biological Chemistry **273**(32): 20196-20204.

Duggleby, R. D. (2006). "Domain Relationships in Thiamine Diphosphate-Dependent Enzymes." Accounts of Chemical Research **39**(8): 550-557.

Ecogene Website - http://ecogene.org/geneinfo.php?eg_id=EG10579.

Emmons, G. T., Campbell, I.M., Bentley, R. (1985). "Vitamin K (Menaquinone) biosynthesis in bacteria: purification and probable structure of an intermediate prior to *o*-succinylbenzoate." Biochemical and Biophysical Research Communications **131**: 956-960.

Emsley, P., Cowtan, K. (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D-Biological Crystallography **60**: 2126-2132.

Fiser, A., Do, R.K., Sali, A. (2000). "Modeling of loops in protein structures." Protein Science **9**: 1753-1773.

Frank, R. A. W., Leeper, F.J., Luisi, B.F. (2007). "Structure, mechanism and catalytic duality of thiamine-dependent enzymes." Cell and Molecular Life Sciences **64**: 892-905.

Fridell, Y.W.C., Villa, J., Attar, E.C., Liu, E.T. (1998). "Gas6 Induces Axl-Mediated Chemotaxis of Vascular Smooth Muscle." Journal of Biological Chemistry **273**: 7123-7126.

Gotz, F., Sedewitz, B. (1991). "Physiological role of pyruvate oxidase in the aerobic metabolism in *Lactobacillus plantarum*." In: Biochemistry and Physiology of Thiamine Diphosphate Dependent Enzymes. Verlag Chemie, Weinheim. 286-293

Gouet, P., Courcelle, E., Stuart, D.I., Metoz, F. (1999). "ESPrpt: multiple sequence alignments in PostScript." Bioinformatics. **15**: 305-308.

Hall, M.O., Prieto, A.L., Obin, M.S., Abrams, T.A., Burgess, B.L., Heeb, M.J., Agnew, B.J. (2001). "Outer Segment Phagocytosis by Cultured Retinal Pigment Epithelial Cells Requires Gas6." Experimental Eye Research **73**: 509-520.

Hall, M.O., Obin, M.S., Prieto, A.L., Burgess, B.L., Abrams, T.A. (2002). "Gas6 Binding to Photoreceptor Outer Segments Requires *gamma*-Carboxyglutamic Acid (Gla) and Ca²⁺ and is Required for OS Phagocytosis by RPE Cells *in vitro*." Experimental Eye Research **75**: 391-400.

Hasson, M. S., Muscate, A., McLeish, M.J., Polovnikova, L.S., Gerlt, J.A., Kenyon, G.L., Petsko, G.A., Ringe, D. (1998). "The Crystal Structure of Benzoylformate Decarboxylase at 1.6 Å Resolution: Diversity of Catalytic Residues in Thiamin Diphosphate-Dependent Enzymes." Biochemistry **37**: 9918-9930.

Hendrickson, W.A., Horton, J.R., LeMaster, D.M. (1990). "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure." The EMBO Journal **9**(5): 1665-1672.

Ho, K. (2007). "Investigation of Catalysis by MenD." MSc. Thesis, University of Saskatchewan, Saskatoon, Canada

Jiang, M., Cao, Y., Guo Z.F., Chen M.J., Chen X.L., Guo Z.H. (2007). "Menaquinone biosynthesis in *Escherichia coli*: Identification of 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate as a novel intermediate and re-evaluation of MenD activity." Biochemistry: **46**(38): 10979-10989.

Jiang, M., Chen, X.L., Guo, Z.F., Cao, Y., Chen, M.J., Guo, Z.H. (2008). "Identification and Characterization of (1R,6R)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate Synthase in the menaquinone Biosynthesis of *Escherichia coli*." Biochemistry: **47**(11): 3426-3434.

Khaleeli, N., Li, R.F., Townsend, C.A. (1999). "Origin of the beta-lactam carbons in clavulanic acid from an unusual thiamine pyrophosphate-mediated reaction." Journal of the American Chemical Society. **121**(39): 9223-9224.

Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallographica D. **52**: 842-857.

Kleywegt, G.J. (1998). "Biomolecular Morphing." Unpublished results.

Koike-Takeshita, A., Koyama, T., Ogura, K. (1997). "Identification of a novel gene cluster participating in menaquinone (vitamin K-2) biosynthesis - Cloning and sequence determination of the 2-heptaprenyl-1,4-naphthoquinone methyltransferase gene of *Bacillus stearothermophilus*." Journal of Biological Chemistry. **272**(19): 12380-12383.

Kulman, J.D., Harris, J.E., Haldeman, B.A., David, E.W. (1997). "Primary Structure and Tissue Distribution of Two Novel Proline-rich *gamma*-Carboxyglutamic Acid Proteins." Proceedings of the National Academy of Sciences of the United States of America. **94**: 9058-9062.

Kulman, J.D., Harris, J.E., Xie, L., David, E.W. (2001). "Identification of Two Novel Transmembrane *gamma*-Carboxyglutamic Acid Proteins Expressed Broadly in Fetal and Adult Tissues." Proceedings of the National Academy of Sciences of the United States of America. **98**: 1370-1375.

Leslie, A.G.W. (1992). Joint CCP3 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26

Lindqvist, Y., Schneider, G., Ermler, U., Sundstrom, M. (1992). "Three-dimensional Structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution." The EMBO Journal **11**: 2373-2379.

Liu, M., Sergienko, E.A., Guo, F., Wang, J., Tittmann, K., Hubner, G., Furey, W., Jordan, F. (2001). "Catalytic acid-base groups in yeast pyruvate decarboxylase. 1. Site directed mutagenesis and steady-state kinetics studies on the enzyme with D28A, H114F, H115F, and E477Q substitutions." Biochemistry. **40**: 7382-7403.

Macova, A. (2005). "Structure-Function Relationship of the Enzyme MenD ((1*R*,6*R*)-2-succinyl-6-hydroxyl-2,4-cyclohexadiene-carboxylate) from *Escherichia coli*." MSc. Thesis, University of Saskatchewan, Saskatoon, Canada

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). "Phaser Crystallographic Software." Journal of Applied Crystallography. **40**: 658-674

Meganathan, R., Bentley, R. (1981). "Biosynthesis of osuccinylbenzoic acid in a men-*Escherichia coli* mutant requires decarboxylation of L-glutamate at the C-1 position." Biochemistry. **20**: 5336-5340.

Meganathan, R., Bentley, R. (1983). "Thiamine pyrophosphate requirement for *o*-succinylbenzoic acid synthesis in *Escherichia coli* and evidence for an intermediate." Journal of Bacteriology **153**: 739-746.

Meganathan, R. (2001). "Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms." Vitamins and Hormones. **61**: 173-218.

Mosbacher, T. G., Mueller, M., Schultz, G.E. (2005). "Structure and mechanism of the ThDP-dependent benzaldehyde lyase from *Pseudomonas fluorescens*." The FEBS Journal **272**: 6067-6076.

- Muller, Y. A., Schumacher, G., Rudolph, R., Schulz, G.E. (1994). "The refined structures of a stabilized mutant and of wild-type pyruvate oxidase from *Lactobacillus plantarum*." Journal of Molecular Biology **237**: 315-335.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997). "Refinement of Macromolecular Structures by the Maximum-Likelihood Method." Acta Crystallographica D, **53**: 240-255
- Navaza, J., Trapani, S. (2008). "AMoRe: Classical and Modern." Acta Crystallographica D. **64**: 11-16.
- Ni, R.Z., Nishikawa, Y., Carr, B.I. (1998). "Cell Growth Inhibition by a Novel Vitamin K is Associated with Induction of Protein Tyrosine Phosphorylation." Journal of Biological Chemistry. **273**(16): 9906-9911
- Palaniappan, C., Sharma, V., Hudspeth, M.E.S. (1992). "Menaquinone (Vitamin-K2) biosynthesis – Evidence that that *Escherichia coli* MenD gene encodes both 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase and alpha-ketoglutarate decarboxylase activities." Journal of Bacteriology. **174**(24): 8111-8118.
- Pang, S. S., Duggleby, R.G., Guddat, L.W. (2002). "Crystal Structure of Yeast Acetohydroxyacid Synthase: A Target for Herbicidal Inhibitors." Journal of Molecular Biology. **317**: 249-262.
- Pang, S. S., Duggleby, R.G., Schowen, R.L., Guddat, L.W. (2004). "The Crystal Structures of *Klebsiella pneumoniae* Acetolactate Synthase with Enzyme-bound cofactor and with an Unusual Intermediate." The Journal of Biological Chemistry **279**(3): 2242-2253.
- Qiagen. (2003). The QIAexpressionist – A Handbook for High-Level Expression and Purification of 6xHis-tagged Proteins. Qiagen Inc. Canada, Mississauga, Ontario.
- Read, R.J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Crystallographica D **278**: 1145-1153
- Rhodes, G. (2000). Crystallography Made Crystal Clear – A Guide of Users of Macromolecular Models. Academic Press, San Diego, California.
- Schenk, G., Leeper, F.J., England, R., Nixon, P.F., Duggleby, R. (1997). The role of His113 and His114 in pyruvate decarboxylase from *Zymomonas mobilis*. European Journal of Biochemistry. **248**: 63-71.
- Schutz, A., Sandalova, T., Ricagno, S., Hubner, G., Konig, S., Schneider, G. (2003). "Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from *Enterobacter cloacae*, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid." European Journal of Biochemistry **270**: 2312-2321.

Sharma, V., Suvarna, K., Meganathan, R. (1992). "Menaquinone (Vitamin-K2) biosynthesis – Nucleotide-sequence and expression of the MenB gene from *Escherichia coli*." Journal of Bacteriology. **174**(15): 5057-5062.

Sharma, V., Meganathan, R., Hudspeth, M.E.S. (1993). "Menaquinone (Vitamin-K2) biosynthesis – Cloning, nucleotide sequence, and expression of the MenC gene from *Escherichia coli*." Journal of Bacteriology. **175**(15): 4917-4921.

Sharma, V., Hudspeth, M.E.S., Meganathan, R. (1996) "Menaquinone (vitamin K-2) biosynthesis: Localization and characterization of the menE gene from *Escherichia coli*." Gene. **168**(1): 43-48

Shi, J., Blundell, T.L., Mizuguchi, K. (2001). "FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties." Journal of Molecular Biology. **310**(1): 243-257.

Shiraki, M., Shiraki, Y., Aoki, C., Miura, M. (2000). "Vitamin K2 (menatetrenone) effectively prevents fractures and sustains lumbar bone mineral density in osteoporosis." Journal of Bone and Mineral Research **15**(3): 515-521.

Suvarna, K., Stevenson, D., Meganathan, R. (1998) "Menaquinone (vitamin K-2) biosynthesis: Localization and characterization of the MenA gene from *Escherichia coli*." Journal of Bacteriology. **180**(10): 2782-2787

Truglio, J. J., Theis, K., Feng, U., Gajda, R., Machutta, C., Tonge, P.J., Kisker, C. (2003). "Crystal Structure of *Mycobacterium tuberculosis* MenB, a key enzyme in Vitamin K₂ biosynthesis." Journal of Biological Chemistry **278**: 42352-42360.

Vermeer, C. (2001). "Role of K Vitamins in the Regulation of Tissue Calcification." Journal of Bone and Mineral Metabolism **19**: 201-206.

Weber, P. (2001). "Vitamin K and Bone Health." Nutrition **17**: 880-887.

Weiss, M. (2001). "Global Indicators of X-ray Data Quality." Journal of Applied Crystallography. **34**: 130-135.