STRUCTURAL INVESTIGATION OF MOSA

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 Biochemistry

University of Saskatchewan

Saskatoon

By

Kurt H. Nienaber

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 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 materials in my thesis.

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

Head of the Department of Biochemistry

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5E5

ABSTRACT

MosA is an enzyme from *Sinorhizobium meliloti* L5-30, a beneficial soil bacterium. Initial investigation into this enzyme categorized it as a methyltransferase. Further investigation revealed that this was incorrect, and that MosA is actually a dihydrodipicolinate synthase, part of the *N*-acetylneuraminate lyase superfamily. One of the characteristics of enzyme superfamilies is their low sequence identity, but relatively high structural similarity. The structural investigation reported here confirms the high structural similarity between MosA and other superfamily members.

Investigation of MosA was carried out by means of x-ray crystallography. It was believed that detailed structural information may shed light into not only the enzymatic mechanism, but also the inhibition of MosA by lysine, the final product of the enzymatic pathway. Insight into enzyme mechanism and inhibition may ultimately prove useful in herbicide or insecticide development, as other dihydrodipicolinate synthases from harmful fungi, bacteria, or plants, make attractive targets for inhibition. Lysine is an essential amino acid for humans, meaning that there is no endogenous lysine production to block the use of these hypothetical inhibitors. Specific inhibitors based on crystal structures have proven to be effective in the past and hopefully, will continue to be useful in the future.

Here we report the structure of MosA, solved to 1.95 Å resolution with lysine 161 forming a Schiff-base adduct with pyruvate. This adduct is consistent with the currently accepted dihydrodipicolinate synthase enzyme mechanism.

ACKNOWLEDGEMENTS

I would like to start by thanking my supervisors, Drs. Louis Delbaere and Fawzy Georges, for providing a wonderful project and immeasurable support and guidance. I also owe a debt of thanks to Dr. David Palmer and Chris Phenix who provided the enzyme characterized in this thesis. I would also like to thank everyone who has helped and guided me through this experience from the Delbaere lab, including Dr. Lata Prasad, Dr. Sanjukta Aich, Julien Cotelesage, Jenny Puttick, Angela Hoffort, and Yvonne Leduc. Without them, I would be lost. Special thanks to my advisory committee members, whose insight and advice is dearly appreciated. I would like to thank the people at the Canadian Light Source for their assistance including Dr. Pawel Grochulski, Dr. Michel Fodje, and Alan Duffy.

For financial support, I would like to thank the Saskatchewan Synchrotron Sciences Institute, the Saskatchewan Canola Development Commision, the College of Medicine, and the College of Graduate Studies and Research.

Most of all I wish to thank my family and friends; Mom, Dad, Ann, Dave, Rebekah, Jared, Jessica, Sam, Mark, Deany, Madison, Rachel, Spencer, Payton, and my Godson Pearson. You guys mean the world to me. Additional thanks to all my aunts, uncles, and cousins. Though too numerous to name, your support is never-ending. And finally to all my friends; Brian, Cam, Rob, Melissa, Bill, Gerald, Shelagh, and cousin Eric, you make the week worth getting through.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DEAE	diethylaminoethyl
DHDPS	dihydrodipicolinate synthase
EDTA	ethylenediaminetetraacetic acid
HMN	2,4,6,7,8,9-hexahydroxy-5-methylcarboxamido nonanoic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_i	disassociation constant for inhibitor binding
K_m	concentration of substrate that produces half-maximal velocity
L-ASA	*L*-aspartate-β-semialdehyde
MOPS	3-[N-Morpholino]propanesulfonic acid
NAL	*N*-acetylneuraminate lyase
PCR	polymerase chain reaction
PEG400	polyethylene glycol; molecular weight of 400
r.m.s.d.	root mean square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TIM	triose phosphate isomerase
Tris	Tris-HCl (hydroxy-methyl)-aminomethane

1.0 Introduction

Proteins serve an important role in the proper functioning of all living organisms. Most enzymes are proteins that catalyze chemical reactions. By reducing the activation energy of a specific chemical reaction, enzymes facilitate chemistry that is vital for an organism's survival. Enzymes are also capable of stereo-specific reactions that would be virtually impossible to achieve in their absence.

In 1958, myoglobin was the first protein structure solved using X-ray crystallography (Kendrew *et al*., 1958). In the following years, it became apparent that solving a protein or enzyme's structure could yield highly useful information. By understanding the mechanism, specifically the residues involved, insight into an enzyme's function and regulation can be gained. This insight can lead to advancements in drug design, development of more economical enzyme reactions for industrial use, or even the creation of new vaccines.

1.1 *N*-acetylneuraminate lyase (NAL) superfamily

An enzyme superfamily has certain characteristics, for example, a low level of sequence identity, but a similar fold with similar active site residues in a conserved orientation (Lawrence *et al*., 1997). In addition, some superfamilies, like the enolase superfamily, share common structural architecture but catalyse a wide range of different reactions (Babbitt and Gerlt, 1997). *N*-acetylneuraminate lyase itself is a class I aldolase

found in numerous organisms that catalyses the cleavage of *N*-acetylneuraminate to *N*-acetyl-D-mannose and pyruvate (Figure 1.1).

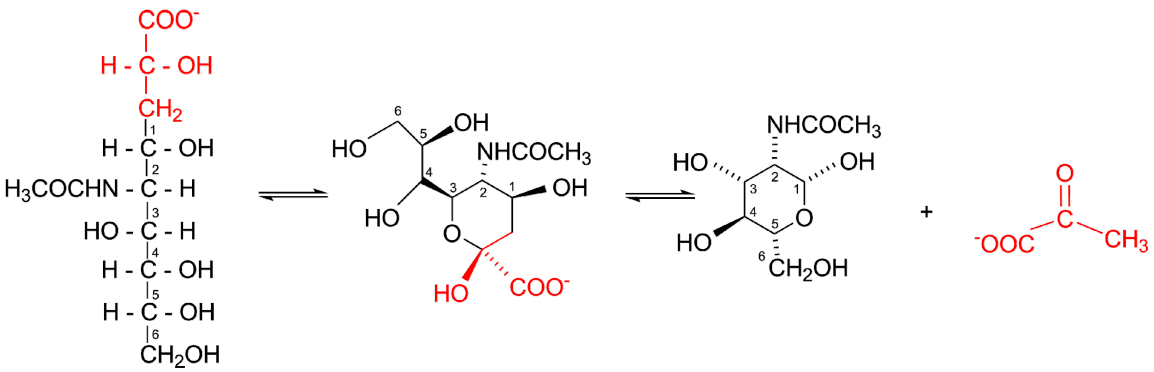
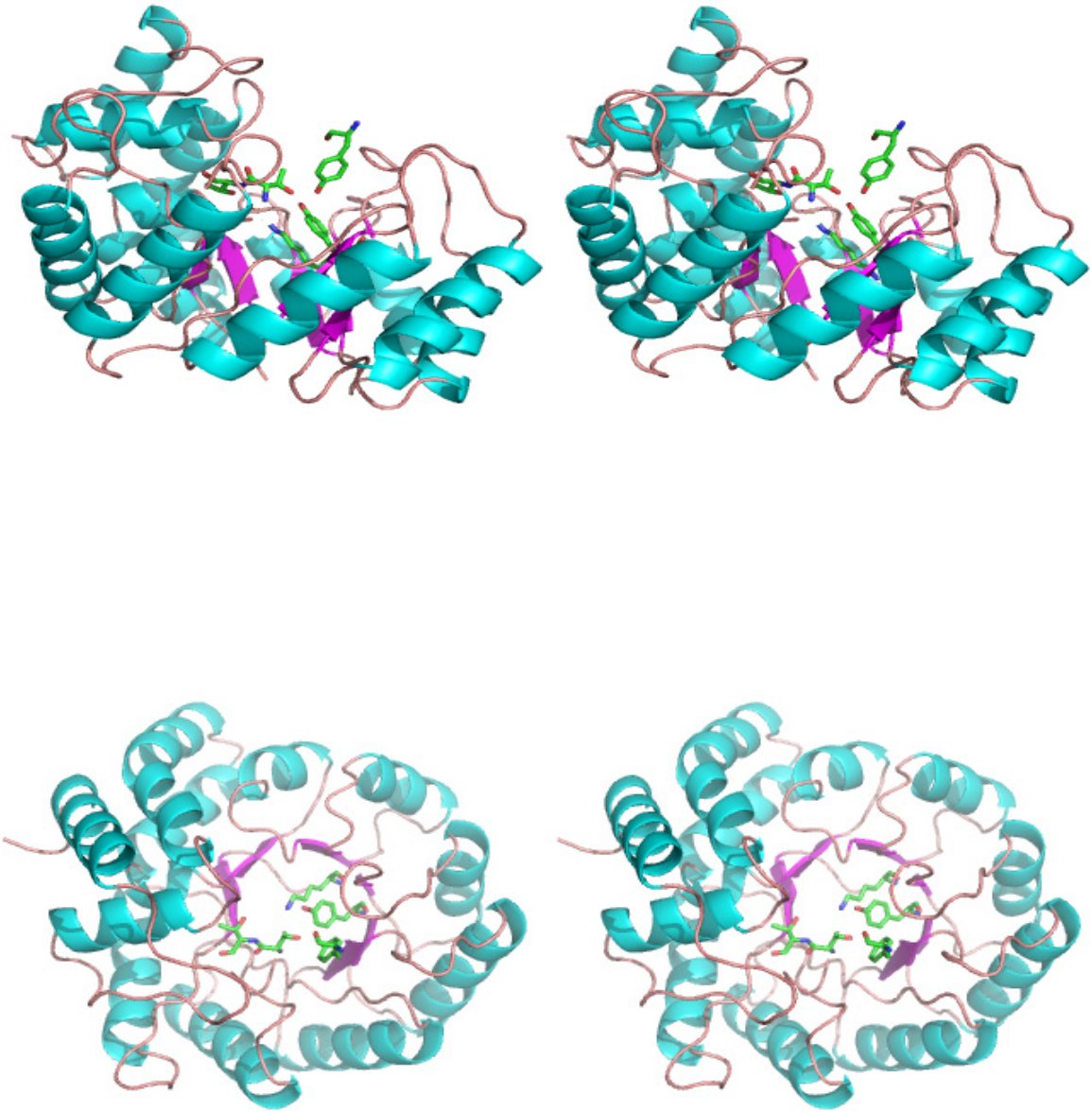


Figure 1.1 The *N*-acetylneuraminate lyase reaction (adapted from Barbosa, 2000). Shown on the left side of the reaction is N-acetyl-neuraminate in linear and ring form. On the right side of the reaction are the two products, N-acetyl-D-mannose and pyruvate.

The NAL superfamily consists of hundreds of enzymes with similar overall structural homology, folding into a triose phosphate isomerase (TIM) barrel (Banner *et al*., 1975) (reviewed by Reardon and Farber, 1995) (Figure 1.2) and a wide range of differing reactions (Figure 1.3). These reactions include the use of an electron sink strategy employed in aldol condensation, dihydrodipicolinate synthase, and dehydratase/ decarboxylase (Babbitt and Gerlt, 1997). These reactions utilize Schiff base formation and in the case of NAL, a Schiff base is formed and a new double bond is produced in the product. Imine or Schiff base formation is a common enzyme intermediate in the NAL superfamily (Lawrence *et al*, 1997) (Figure 1.3).



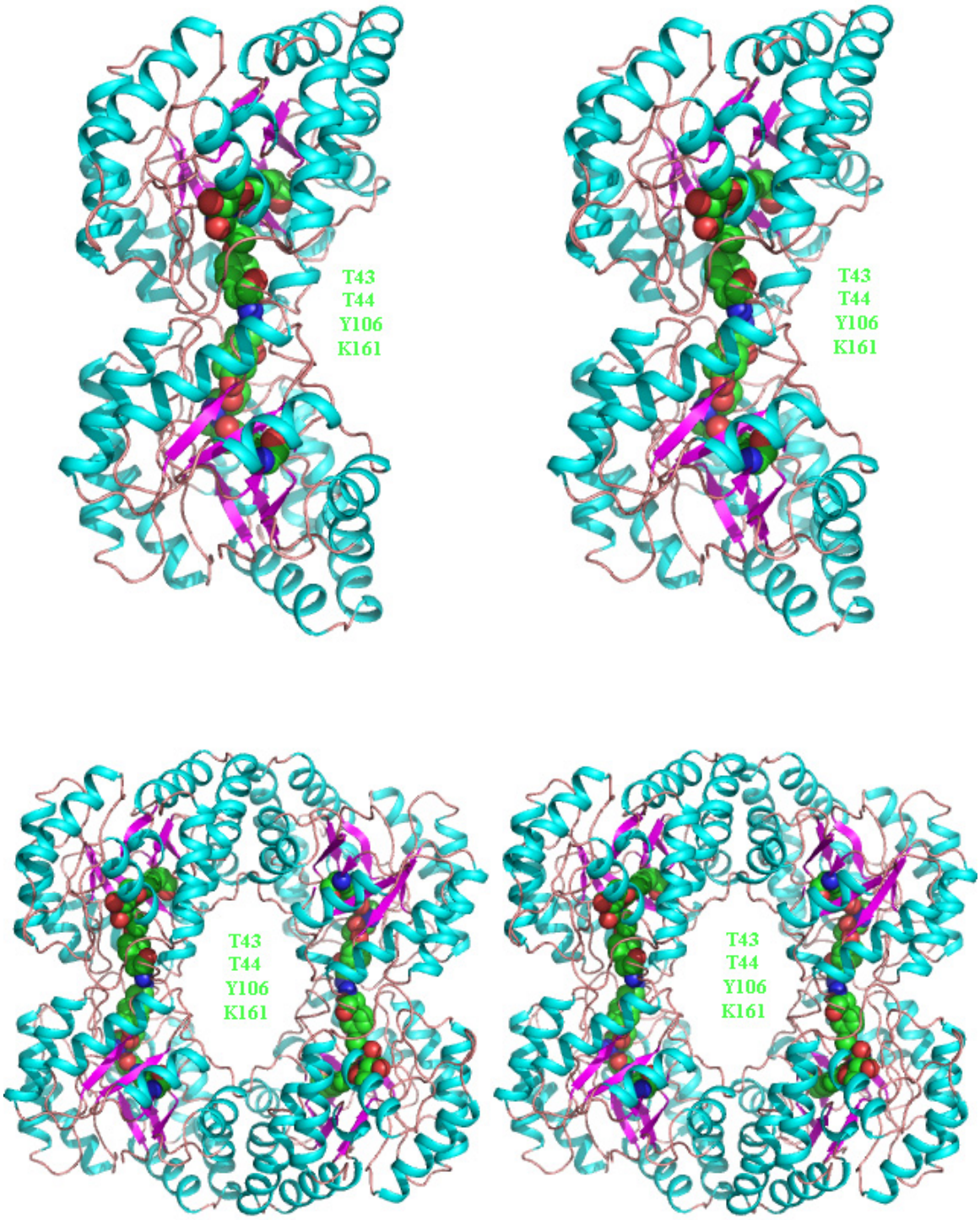


Figure 1.2 Overall *E. coli* DHDPS folding (1DHP). Above, the single subunit, two interconnected subunits, and homotetramer show similar folding to other known NAL family members. The $(\beta/\alpha)_8$ TIM barrel motif, characteristic of the NAL superfamily, is

evident. The outer 8 α helices wrap around 8 parallel β strands, orienting the key active site residues pictured in green.

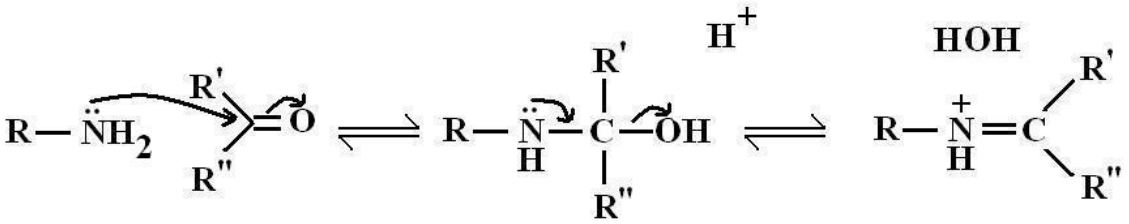


Figure 1.3 Generic Schiff base formation reaction. A carbonyl group undergoes nucleophilic attack from an amino group nitrogen followed by a dehydration facilitated by a nearby acid.

1.1.1 Dihydrodipicolinate synthases (DHDPSs)

DHDPSs are aldolases catalyzing the condensation of *L*-aspartate-β-semialdehyde (*L*-ASA) to pyruvate and dihydrodipicolinic acid. In a ping-pong mechanism, DHDPS utilizes an imine intermediate in the reaction (Figure 1.3). There are currently twelve DHDPS structures reported in the Protein Data Bank, originating from five different organisms. First solved in 1995, DHDPS from *Escherichia coli* (*E. coli*) (1DHP) (Mirwaldt *et al*., 1995), shares a 45% identity with MosA, the DHDPS from *Sinorhizobium meliloti* (*S. meliloti*). Seven DHDPS structures deposited are from *E. coli*, six of them being mutants that reduce enzymatic activity (1S5T, 1S5V, 1S5W, Dobson *et al*., 2004a; 2A6L, 2A6N, Dobson *et al*., 2005a; 2OJP). T44V, Y107F, Y133F mutants showed a marked reduction in activity, confirming the notion of a catalytic triad made up of V44, Y107, and Y133 (Dobson *et al*., 2004a). The exact function of R138 in the enzyme mechanism is unknown, but it is suspected that R138 may play a role in *L*-ASA

binding. To examine the possible role of R138, two mutants were constructed, R138H and R138A, both showing reduced activity and lending credence to the theory that R138 does play a vital role in stabilizing the catalytic triad (Dobson et al., 2005a). The question of *L*-ASA binding, and the possible role R138 would play, remains unanswered.

There are two non-mutated *E. coli* DHDPS structures that have been deposited, the native (1DHP), and a ligand bound structure containing two lysine amino acids bound at the allosteric site (1YXD)(Dobson *et al*., 2005b). As discussed in Section 1.1.3, DHDPS from *E. coli* and MosA from *S. meliloti* are involved in the biosynthesis of lysine. As such, lysine, as the end product, is a negative feedback inhibitor of DHDPS from *E. coli* and MosA from *S. meliloti*.

All of the reported structures share an overall structural similarity. A multiple sequence alignment shows the overall identity and homology between MosA and other DHDPSs, the highest sequence identity being 45% with 1DHP, the *E. coli* DHDPS (Figure 1.4). The structure of MosA, the DHDPS from *S. meliloti* reported here, also shares a high degree of structural homology, confirming its place in the DHDPS family. As seen in Figure 1.5, structural overlays between currently solved DHDPSs and *E. coli* DHDPS show similarities, including the overall TIM barrel motif. Further discussion on *E. coli* DHDPS and MosA is found in Section 4.2.1. NAL from *Haemophilus influenzae* (*H. influenzae*) was solved in 2000 with three inhibitors bound (Barbosa *et al*., 2000). Comparisons between MosA and this NAL are discussed in Section 4.2.2.; however, as seen in Figure 1.5, a structural overlay of the Cα atoms of *E. coli* DHDPS and *H. influenzae* NAL reveals some differences in helix positioning, but overall structural similarity. Of further interest in the *H. influenzae* NAL structure is the presence of three

different inhibitor molecules, bound in separate structures. One inhibitor is sialic acid alditol (2,4,6,7,8,9-hexahydroxy-5-methyl-methylcarboxamido nonanoic acid: HMN), which binds directly to the active site and provides some insight into possible mechanisms of inhibition (Figure 1.6)(Figure 4.9, 4.10)(pg. 58,59). In addition to *E. coli* and *H. influenzae*, structures of DHDPS from *Thermatoga maratima* (*T. maratima*) (1O5K) (Figure 1.5); DHDPS from *Mycobacterium tuberculosis* (*M. tuberculosis*) (1XXX) (Figure 1.5) (Kefala *et al*., 2008); and DHDPS from *Aquifex aeolicus* (*A. aeolicus*) (2EHH) (Figure 1.5), have been solved. Compared with the structure of MosA discussed in this thesis, all show folding similarities, and would seem to confirm MosA as a member of this enzyme family.

Initial investigation into MosA function resulted in a sequence being published which was somewhat erroneous, with a frame shift causing forty C-terminal amino acids to be removed (Rao *et al*., 1995) (Tam *et al*., 2004). Complementation and deletion experiments identified MosA as a methyltransferase (Rao *et al*., 1995). This seemed to defy conventional thinking about enzyme superfamilies, as MosA belonged to the NAL superfamily. While sequence may vary in a superfamily, the overall set of reactions tends to be conserved. No other NAL superfamily member is a methyltransferase. Work in the Palmer laboratory revealed the original error in the reported sequence. The new sequence was deposited and MosA enzyme activity was assayed. After confirming aldolase activity, further work with isothermal titration calorimetry revealed no interaction between MosA and its supposed methyltransferase substrates (Phenix, 2007).

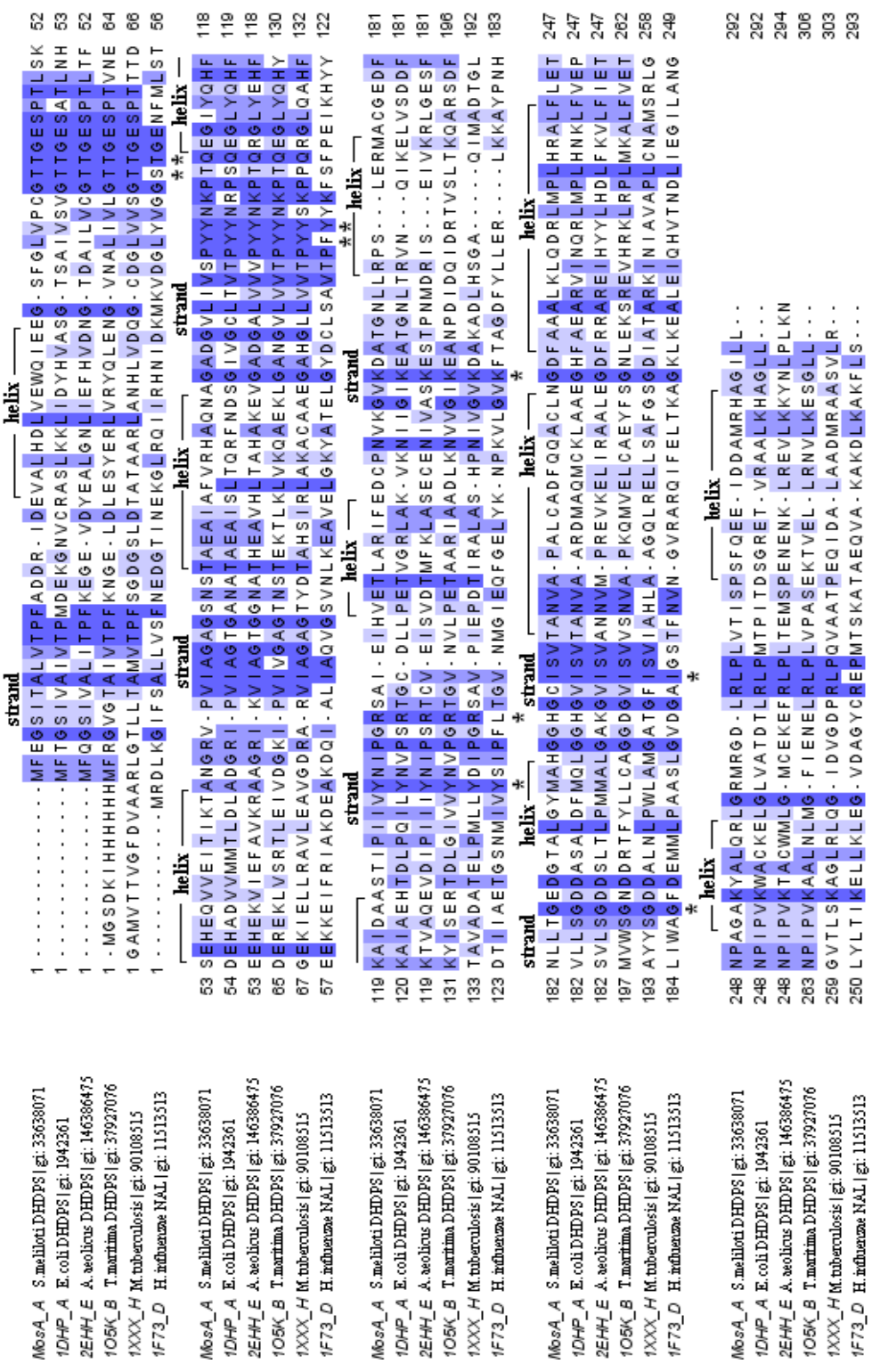


Figure 1.4 Sequence alignment of MosA, the DHDPS from *S. meliloti*, and other NAL family members for which the structure is known. The alignment was generated using Clustalw2 with the gonnet scoring matrix (gap penalty 10, gap extension penalty 0.2), and visualized with Jalview (Clamp *et al*., 2004). Identical and homologous residues

shown in blue, darkest blue identity, lighter blue showing similarity. Active site residues denoted with an asterisk.

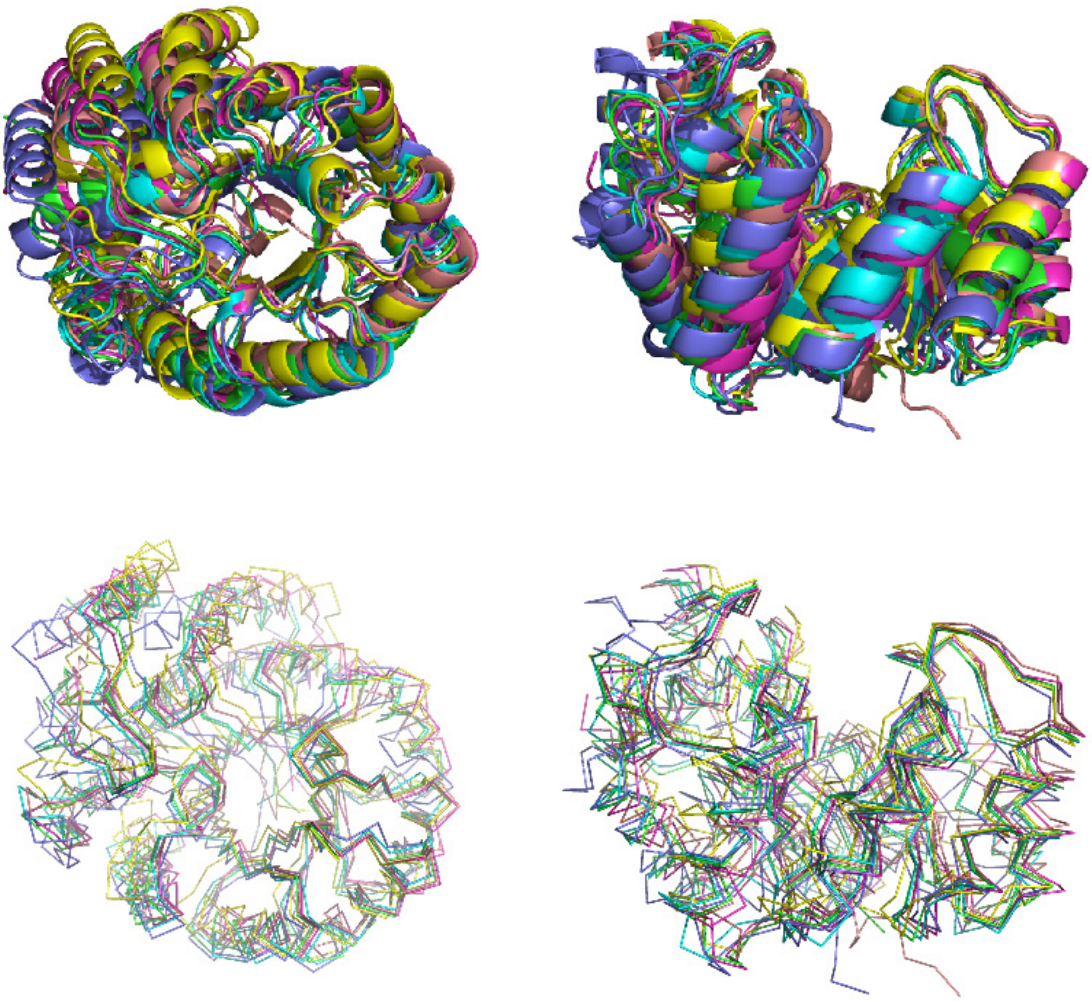


Figure 1.5 Multiple overlay of all currently solved NAL/DHDPS family members. Shown from the top and side, the TIM barrel motif is evident in all of the enzymes from the NAL superfamily currently solved. Pictured are MosA (green), *E. coli* DHDPS (1DHP; light blue), *A. aeolicus* DHDPS (2EHH; magenta), *T. maritima* DHDPS (1O5K; yellow), *M. tuberculosis* DHDPS (1XXX; peach), and *H. influenzae* NAL (1F73; dark blue). Final r.m.s.d values ranged from 1.2 – 2.3 Å and can be found with the residue ranges used in Table 3.5.

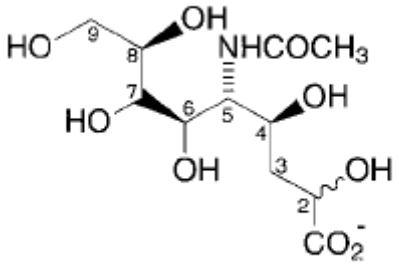


Figure 1.6 Diagram of sialic acid alditol (HMN).

1.1.2 MosA

MosA is the DHDPS found in *S. meliloti* (Dixon, 1969). It was originally misidentified as a methyltransferase (Rao *et al*., 1995) but research done in recent years has confirmed MosA is indeed a DHDPS. The MosA subunit has a molecular mass of around 33.4 kDa and, like other members of the NAL superfamily, has a classic TIM barrel tertiary structure. MosA aldolase activity was confirmed with a K_m of around 0.27 mM for pyruvate and a 0.13 mM for *L*-ASA (Tam *et al*., 2004).

The proposed mechanism of action involves a Schiff base intermediate, like other members of the NAL superfamily (Babbitt and Gerlt, 1997). Key active site residues are T43, T44, Y106, Y132, R137, K161, G186 and I203 (*S. meliloti* numbering). As seen in Figure 1.7, the active site residues centre around K161, as pyruvate forms an imine with the lysine side chain, completing the first step of the reaction. Reported in this thesis, the structure of MosA with pyruvate bound is the first such structure deposited in the Protein Data Bank. This imine intermediate has been documented (Borthwick *et al*., 1995; Laber *et al*., 1992), but has never been captured before via protein crystallography. This thesis reports the imine of pyruvate and lysine bound at the active site, which tends to confirm

the currently accepted mechanism of action (Figure 1.8) and supports the role of MosA in the lysine biosynthetic pathway as a DHDPS. Inhibition of DHDPSs by lysine has been documented (Dobson et al., 2004b; Yugari and Gilvarg, 1965), and was proven to occur at an allosteric site (Figure 1.11) (Figure 4.9; pg. 57). This allosteric site is located at the outside edge of the *E. coli* dimer; however, the mechanism of lysine inhibition via binding to the allosteric site remains unknown.

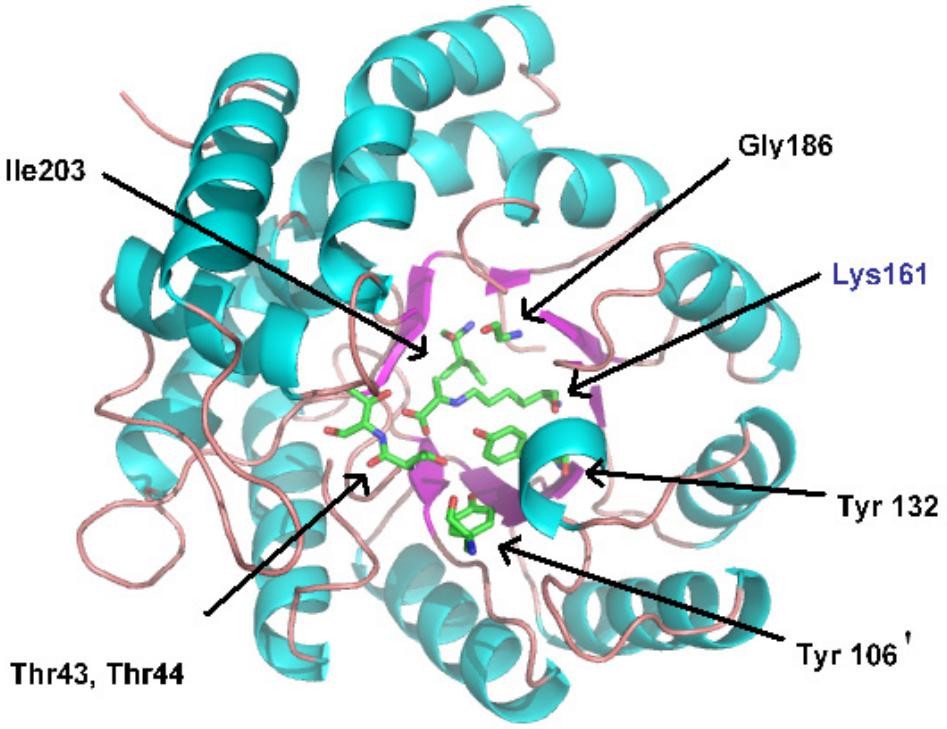


Figure 1.7 View of the MosA active site. Key active site residue carbons are shown in green, oxygens red, and nitrogen blue.

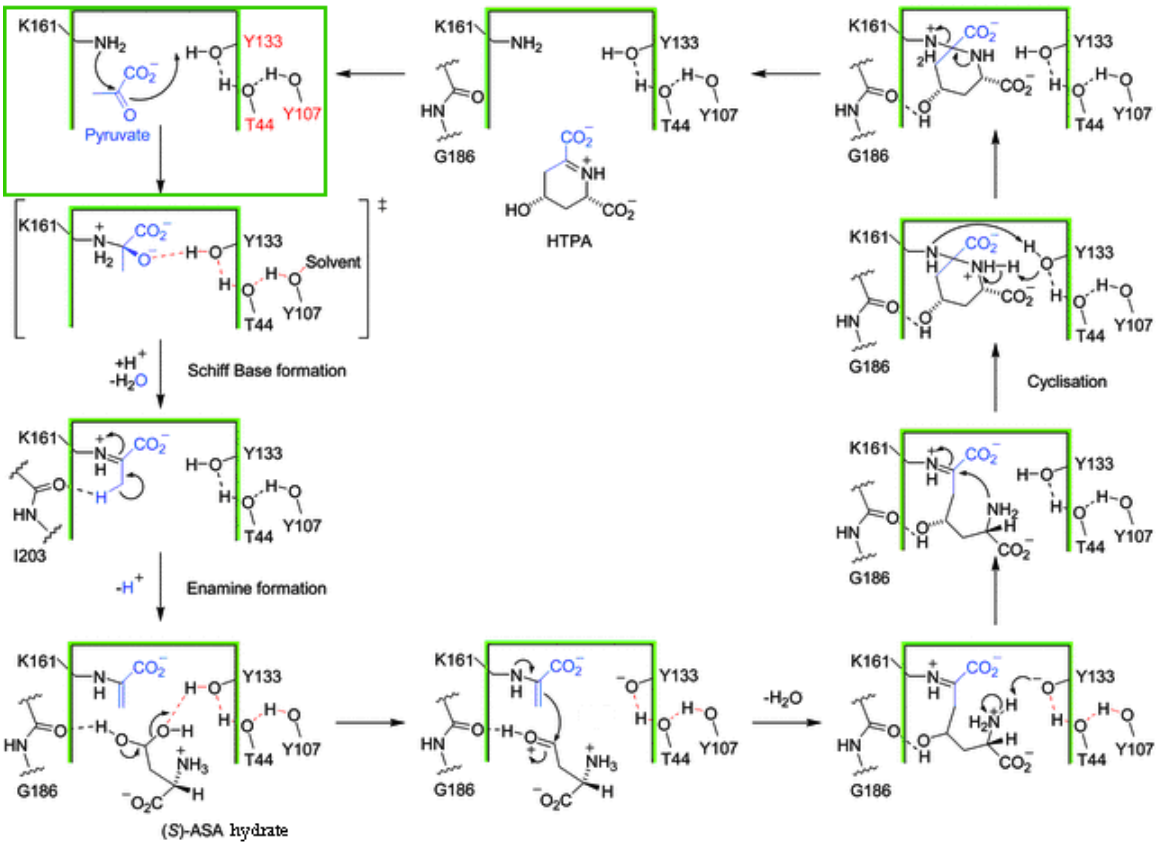


Figure 1.8 The accepted mechanism of DHDPS enzyme activity. Pyruvate binds via nucleophilic attack by the Nε on K161, highlighted in the green box. A Schiff base is formed following a dehydration step with Y133 (*E. coli* numbering) forming a catalytic triad with T44 and Y107; acting first as a Lewis acid to stabilize the carbonyl oxygen intermediate, then later as a Brønsted-Lowry acid in the activation of *L*-ASA. Enamine formation is the end result of pyruvate binding and the activation of a methyl carbon hydrogen by the carbonyl oxygen of I203 with the proton picked up by solution. With the pyruvate covalently bound to K161, *L*-ASA binds via hydrogen bonds with G186 and Y133. Following carbonyl formation, an attack from the methylene electrons binds the pyruvate and *L*-ASA together, with the carbonyl oxygen of G186 acting as a Lewis base. Y133 transfers a proton to the amino group of *L*-ASA, which then attacks the Shiff base complex, completing the cyclization. Y133 is regenerated with a proton from Nε, which itself is regenerated following detachment of (4S)-4-hydroxy-2,3,4,5-tetrahydro-(2S)-dipicolinic acid (HTPA) (adapted from Dobson *et al*., 2005b).

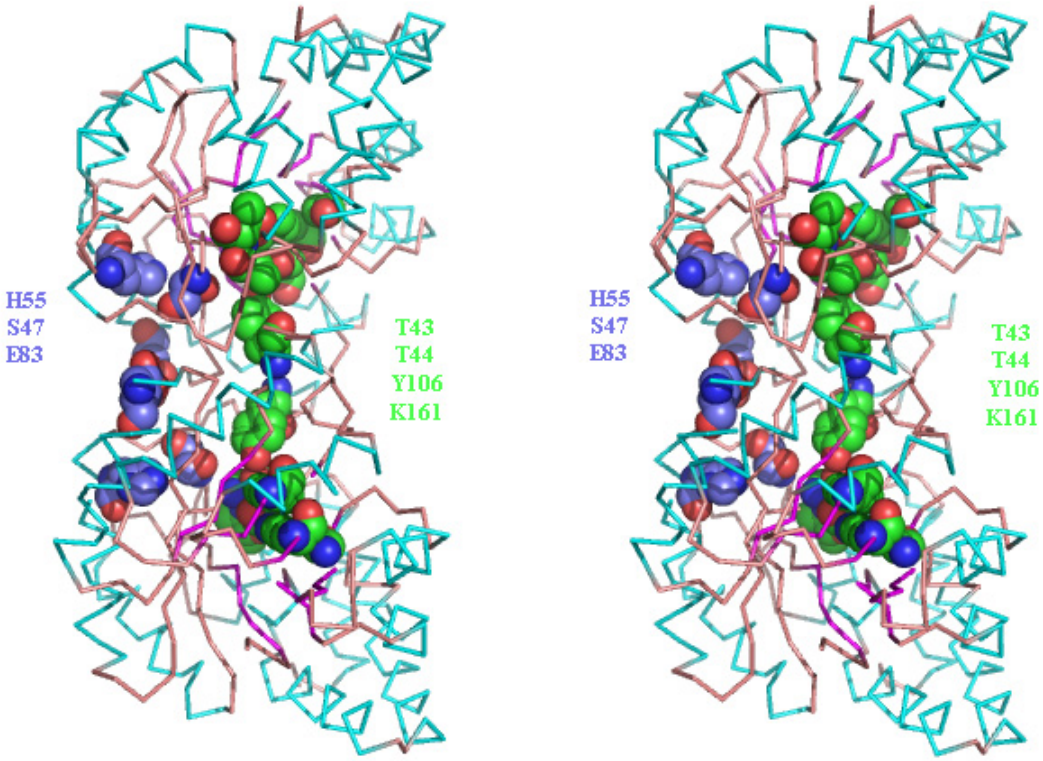


Figure 1.9 Stereo view of the *E. coli* DHDPS dimer. Active site residue carbons are pictured in green, allosteric site residues in light blue. Located on the exterior of the enzyme (right side of the picture), the allosteric site is located more than 5 Å from the active site (center of the molecule). Nitrogens are pictured in dark blue, oxygens in red.

1.1.3 Lysine biosynthesis

The production of lysine in any organism is vital to its survival. Lysine is an essential amino acid for humans, resulting in a need to obtain lysine from primarily dietary sources. In the case of *S. meliloti*, MosA is inhibited by lysine with a K_i of about 0.7 mM (Phenix, 2007). Binding of lysine to MosA complexed with pyruvate is cooperative, with the second lysine binding giving a large exothermic enthalpy value (Phenix, 2007). Lysine binding was also shown to be non-competitive with respect to pyruvate, supporting earlier work done by Dobson (Dobson *et al*., 2005b). Inhibition by

lysine is discussed in Section 4.2.1.1, but shown in Figure 1.9; binding occurs at an allosteric site in *E. coli* DHDPS. This site is located at the outer edge of the *E. coli* DHDPS dimer, and thus the outer edge of the tetramer, but contrary to earlier reports, no channel or direct connection can be observed (Dobson *et al*., 2005b).

DHDPS is found in the diaminopimelate pathway, which is the lysine-producing pathway in lower fungi, bacteria, and green plants (Scapin and Blanchard, 1998). It is for this reason that inhibition of key enzymes is increasingly attractive to researchers looking for novel antibiotics and herbicides (Coulter *et al*., 1999; Cox *et al*., 2000; Hutton *et al*., 2003). Humans have no endogenous lysine production, and thus do not have a DHDPS that would be inhibited by an antibiotic or herbicide developed in this fashion. While *S. meliloti* is a beneficial soil bacterium, structural similarities between DHDPSs make investigation into this particular enzyme valuable (Dixon, 1969).

1.1.4 Objectives

After work in Palmer laboratory revealed that MosA was a member of the DHDPS family of enzymes, it was clear that further information was needed. Structural investigation of MosA was undertaken to provide answers into mechanism of action, and inhibition by lysine. Answers to the role of R137 in enzyme activity and *L*-ASA binding were also sought out, given the current debate in the literature. Given that DHDPS family members make attractive drug targets, further information about the DHDPS family, including MosA, would prove valuable.

1.2 Methods of X-ray Crystallography

Crystals are regularly repeating units of identical shape and size. They can be categorized into fourteen Bravais lattices. In crystals, the smallest repeating unit that arranges in these space groups is called an asymmetric unit. This regular packing is what allows researchers to gain information about the protein by placing the crystal in a high-energy X-ray beam (>12 KeV). These X-rays are generated either by a rotating anode source, in the researcher's lab, or by a synchrotron. The latter of these produces X-rays of very high intensity and low divergence. The theory is that electrons around the atoms absorb the electromagnetic radiation, and then re-emit that radiation, at the same energy along a new vector. Because crystals are essentially a three-dimensional diffraction grating, a diffraction pattern can be used to determine information about the unit cell in that crystal, including the shape of the electron clouds that diffracted those X-rays i.e. the protein structure. A space group is the three-dimensional arrangement of the molecules in the unit cell. By combining the Bravais lattices with all the point groups, and allowing symmetry elements such as translations, screw axes, and glide planes, there are 230 possible space groups but because proteins molecules are chiral, they can only crystallize in 65 possible space groups.

1.2.1 Diffraction

Diffraction of X-rays occurs when electrons surrounding atoms in a molecule absorb and re-emit X-rays. W. L. Bragg showed that X-ray reflections could be observed using specific incident angles and wavelengths when a crystal is placed in front of them (Bragg, W. H. and Bragg, W. L., 1913). Furthermore, Bragg showed that modeling a set of

parallel planes separated by a set perpendicular distance d, a peak of X-ray intensity would be observed if the reflection off those planes interfered constructively. A reflected beam will occur at an incident angle θ and a reflected angle θ with respect to the planes, and at an angle 2θ from the direct beam as described in Bragg's Law (2.1).

$$2d_{hkl}\sin\theta = n\lambda \qquad 2.1$$

The indices *h*, *k*, and *l* refer to the number of planes in the set per unit cell, in the *x*, *y*, and *z* directions respectively. The interplanar spacing is thus d_{hkl} and the x-ray wavelength λ, and n is an integer, 1, 2, 3 …. The resulting diffraction pattern is a result of all the atomic scattering contributions. The scattering observed is dependent on the location and type of each atom in the unit cell. Each atom's ability to scatter X-rays can be described by the equation 2.2.

$$f = \int_r \rho(\mathrm{r}) e^{(2\pi i \mathrm{r} \bullet \mathrm{S})} dr \qquad 2.2$$

The scattering factor then is dependent on ρ(r), the electron density at position r, and |S|, where $|S| = (2\sin\theta)/\lambda$. The vector summations of all the scattering are called the structure factors ($F(S)$).

$$F(S) = \sum_{j=1}^{n} f_j e \exp(2\pi i \mathrm{r}_j \bullet \mathrm{S}) \qquad 2.3$$

The structure factor depends on atom type and location. The intensity of the diffracted X-rays measured is proportional to the square of the structure factor amplitudes. A Fourier transform can be used to describe the relationship between reciprocal space and real space. A crystal's diffraction pattern is a Fourier transform of the crystal, and more specifically, of its electron density. The electron density at position *x*, *y*, *z* [ρ(*xyz*)] in the crystal is thus the Fourier transform of the structure factors $F(hkl)$.

$$\rho(xyz) = \frac{1}{V}\int_h\int_k\int_l F(hkl)\exp[-2\pi i(hx+ky+lz)]dhdkdl \qquad 2.4$$

Because data are produced at distinct locations, the integral can be reduced to a Fourier summation.

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l |F(hkl)|\exp[-2\pi i(hx+ky+lz)+i\alpha(hkl)] \qquad 2.5$$

V is the volume of the unit cell and each structure factor, $F(hkl)$, describes a specific reflection in the diffraction pattern. Each term in the above summation is a three-dimensional wave with frequency *h* in the *x* direction, *k* in the *y* direction, and *l* in the *z* direction. These waves, different with each *hkl* value, have amplitude F_{hkl} and phase α_{hkl}. Intensities are measured during an experiment. These intensities are proportional to $|F_{hkl}|$. There is no lens to focus diffracted X-rays, and thus no way to obtain the phase angles directly through experimental data collection. The result is that other means of estimating phases (α_{hkl}) must be employed. This is known as the phase problem in crystallography.

A set of parallel planes in the crystal will diffract identically giving rise to one reflection. The crystal is placed at the centre of the Ewald Sphere of Reflection, and since reflections observed satisfy the Bragg equation, they represent where the reciprocal lattice points touch the sphere (Figure 1.10). Therefore, examination of the diffraction pattern observed can yield information about the crystal lattice.

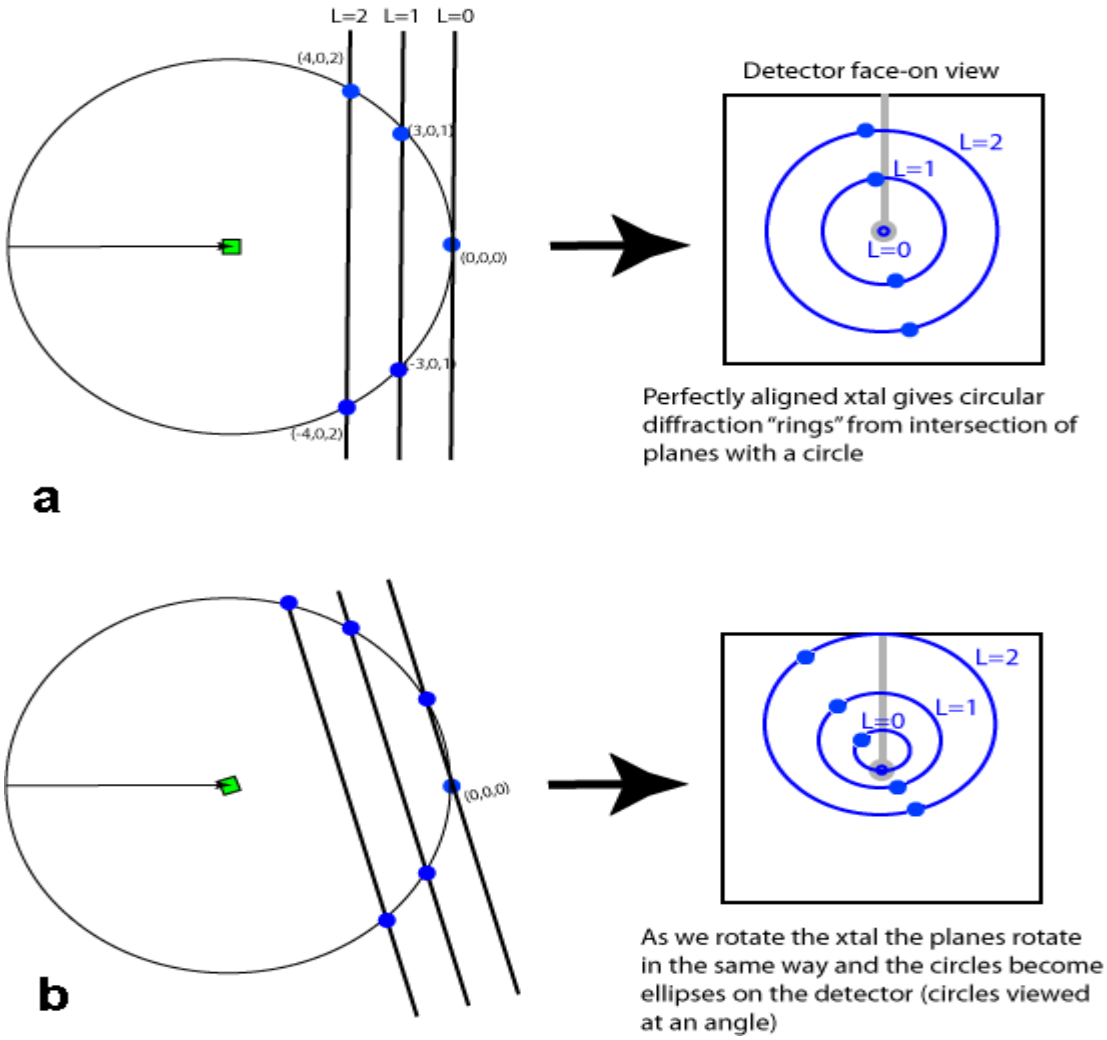


Figure 1.10 Diagram illustrating general indices of reflections as they intercept the Ewald Sphere. By rotating the crystal during experimentation, different lattice points intercept the sphere, satisfying Bragg's Law, and producing a reflection. Depicted are the views of these lattice points as they would appear on an X-ray detector (adapted from Jeffrey, 2006).

1.2.2 Solution of Structure

1.2.2.1 Molecular Replacement

The most common method of solving the phase problem of protein crystals is called molecular replacement. By using phase information from a previously solved protein structure, one can approximate the phases for the protein of interest. Provided an amino acid identity of at least 25% to the protein of interest is present (McCoy *et al.*, 1997), the relative atomic positions of the search model should provide enough information to solve the structure of the protein of interest. To do this, the Patterson function ($P(uvw)$) of each molecule is compared.

$$P(uvw) = \frac{1}{\mathrm{V}} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] \qquad 2.6$$

The Patterson function is the summation of the square of the structure factor amplitudes with all phase angles equal to zero. Due to the fact that all phase angles of a Patterson function are set to zero, it is possible to calculate a Patterson function for an unknown protein, with no prior phase information needed. Each peak in the Patterson function is the product of two atomic scattering factors separated by the vector (u,v,w). In simpler terms, the Patterson function is a pairwise sum of interatomic vectors, each corresponding to the electron density of the two contributing atoms. A Patterson peak thus corresponds to a vector between two atoms within the unit cell. Small distances between atoms in the same protein molecule will create small vector distances, and are called self-Patterson vectors. These intramolecular vectors are used in determining the correct rotation parameters from the search model which are applied to the protein of interest. Large intermolecular distances between atoms will create large vector distances,

called cross-Patterson vectors, and are used in the determination of the correct translation parameters.

A Patterson function is calculated for the starting model and then for a large number of different orientations of that same model. The best solution will have a good overlap between the search model's Patterson function, and the Patterson function of the unknown protein. Using self-Patterson vectors, the correct rotation is calculated first. The second step is to use the cross-Patterson vectors within the unit cell to determine the correct translation. Obtaining calculated phases from the newly rotated and translated model allows an electron density map to be calculated for the unit cell of the protein of interest. An electron density map is calculated by performing a Fourier transform on the structure amplitudes and calculated phases.

1.2.2.2 Refinement of Structure

Once an electron density map is obtained, the researcher must adjust the model of the protein into the density using three-dimensional modeling software. Initially, it may not be possible to see electron density for all of the atoms, side chains, carbonyl oxygens, or even continuous density along the α-carbon backbone. Because data are usually cut off at around 2.5 Å for initial molecular replacement, side chain density may not always be present. If the data are not limited in resolution, a molecular replacement solution may not be found. Should the electron density not be continuous, the crystallographer must modify the electron density. This can be accomplished by several techniques, mostly by a combination of them all.

The model can be adjusted into as much electron density as is possible, and then new phases calculated from those improved atomic positions. The original phases may be combined with the calculated phases to improve the starting point, and allow the model to be readjusted into the density generated from the new combined phases. If the model fits closer to the actual structure, the phases generated from it will also improve, allowing the cycle to continue. A figure of merit may be assigned to the phases based on the cosine of the phase error. As the error gets smaller, the figure of merit approaches a value of 1. This number serves to monitor the phase combination process, and is a useful tool in solving the structure.

2.0 Materials and Methods

2.1 Reagents, Supplies, and Equipment

Reagents, supplies, and equipment used in the described experiments are listed in Table 2.1. The commercial suppliers and their addresses are listed in Table 2.2.

2.2 Crystal Structure of MosA complexed with pyruvate

2.2.1 Protein Purification

The *mosA* gene was cloned, by Chris Phenix in the Palmer laboratory, by PCR-amplification and sub-cloned into a pET-28b expression vector containing an N-terminal poly-histidine tag (Tam *et al.*, 2004). Recombinant protein was expressed in *E. coli* BL-21(DE3) and pelleted cells were lysed in a buffer containing 50 mM Tris-HCl (hydroxy-methyl)-aminomethane (Tris buffer), pH 8.0, 10 mM imidazole, 500 mM NaCl, and 12.5% v/v glycerol. A Chelating Sepharose FF column charged with Ni^{2+} and an EDTA gradient were used to purify MosA from native proteins. Dialysis against 50 mM Tris-HCl, pH 8.0, and a DEAE-Sepharose FF column with an NaCl gradient were used to further purify MosA.

Enzyme activity was assayed (as described by Borthwick *et al*., 1995) and aspartate ß-semialdehyde (ASA) was synthesized by the method of Roberts (Roberts *et al.*, 2003). MosA was then frozen in solution at –80˚C with a 10% v/v glycerol solution used as cryoprotectant.

Table 2.1 Table of biological and chemical reagents, supplies, and equipment.

Item	Supplier
Chemical Reagents	
Ammonium Sulfate	EM Science
Glycerol	Fisher Scientific
Imidazole	Fluka
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Sigma
Lysine	Sigma
2-mercaptoethanol	Sigma
3-[N-Morpholino]propanesulfonic acid (MOPS)	Sigma
Nickel-nitrilotriacetic acid (Ni-NTA)	Qiagen
Nitrogen (liquid)	Praxair
Luria broth	Sigma-Aldrich
Polyethylene Glycol 400	Fluka
Potassium Carbonate (K_2CO_3)	Sigma
Potassium Chloride (KCl)	BDH
Pyruvate	Sigma
Sodium Chloride (NaCl)	BioShop
Sodium Phosphate Monobasic (NaH_2PO_4)	BioShop
Tris-HCl (hydroxy-methyl)-aminomethane (Tris)	BioShop
Supplies and Equipment	
Bench-model pH meter, Corning 125	Dow Corning
Chelating Sepharose FF column	AP Biosciences
Cryoloops	Hampton
DEAE-Sepharose FF column	AP Biosciences
Dow Corning High Vacuum Grease	VWR
DynaPro-MS800	Proterion

Falcon Tubes	VWR
Microcentrifuge tubes	VWR
Micro Cover Glass (No. 1, 22 mm square)	VWR
NAP-25 columns	Pfizer-Pharmacia
Silicon Graphics Indigo^2 computer	SGI
24-Well VDX plates	Hampton
X8 Proteum	Bruker AXS

Table 2.2 Table of names and addresses of suppliers.

Company	Address
AP Bioscience	Amersham Bioscience, Piscataway, NJ, USA
BDH	British Drug House, Saskatoon, SK, Canada
BioShop	BioShop Canada Inc., Burlington, ON, Canada
Bruker AXS	Bruker AXS, Inc., Madison, WI, USA
Dow Corning	Dow Corning Corp., Midland, MI, USA
EM Science	EM Industries Inc., Gibbstown, NJ, USA
Fisher Scientific	Fisher Scientific Company Canada, Ottawa, ON, Canada
Fluka	Sigma-Aldrich Canada Ltd., Oakville, ON, Canada
Hampton	Hampton Research, Lugana Niguel, CA, USA
Pfizer-Pharmacia	Pfizer Inc., New York, NY, USA
Praxair	Praxair, Saskatoon, SK, Canada
Proterion	Wyatt Technology Corp, Santa Barbara, CA, USA
SGI	Silicon Graphics Inc., Mountain View, CA, USA
Sigma	Sigma Chemical Co., St. Louis, MO, USA
Sigma-Aldrich	Sigma-Aldrich Canada Ltd., Oakville, ON, Canada
VWR	VWR, Mississauga, ON, Canada

2.2.2 Crystallization conditions

MosA protein, purified as described in Section 2.2.1, was screened using Hampton's Screen Kit 2 (Jancarik and Kim, 1991; Cudney *et al*., 1994). Final crystallization conditions were optimized from the most promising results of this screen. Orthorhombic MosA crystals used in diffraction were grown using an optimized well solution of 2 M ammonium sulfate, 100 mM Tris buffer pH = 8.0, and 2% v/v polyethyleneglycol 400 (PEG400). The well solution was then mixed with the protein solution in a 1:1 ratio and crystals grew overnight at room temperature.

Trigonal MosA crystals were grown in 2 M ammonium sulfate, 100 mM Tris buffer pH = 8.0, and 2% v/v PEG400. The well solution was mixed in a 1:1 ratio with the protein solution and crystals grew overnight at room temperature.

2.2.3 Cryoprotection and soaking of crystal samples

Crystals were harvested and soaked at room temperature in a cryoprotectant solution consisting of 2 M ammonium sulfate, 100 mM Tris buffer, 2% v/v PEG400, 10% glycerol, 100 mM pyruvate, and 100 mM lysine. Crystals were soaked for 10 minutes and then flash cooled in liquid nitrogen. Trigonal MosA crystals were previously grown using these same conditions as described in Section 2.2.2, and were soaked in the same manner as orthorhombic crystals, then diffracted as described in Section 2.2.4.

2.2.4 Data Collection and Processing

X-ray data were collected at the Canadian Light Source beamline 08ID-1 for orthorhombic MosA crystals soaked with pyruvate and lysine. Data were collected in a

nitrogen stream at 105 K using a MAR225 CCD detector. A total of 360 images were collected with 0.5 degree oscillation per image around the omega axis. A wavelength of 1.33 Å was used. The data were indexed using XDS and then integrated using Mosflm (Leslie, 1992). The data were finally scaled using XDS (Kabsch, 1993) (Table 2.3).

MosA crystals that grew in a trigonal space group were taken to the Advanced Photon Source in Illinois, and were diffracted on BioCARS beamline 14BMC using an ADSC detector. A total of 360 images were collected with an oscillation of 1˚ per image around the phi axis. A wavelength of 0.9 Å was used. Data were finally integrated and scaled using HKL2000 (Otwinowski, Z. and Minor, W., 1997) (Table 2.3).

2.2.5 Molecular Replacement

Molecular replacement was the method chosen to solve the structure of MosA from this data set as previous structures of DHDPSs had been solved with a sequence identity of 45%. Molecular replacement was carried out using Phaser (McCoy *et al.*, 2007). Final molecular replacement solutions can be found in Table 3.1 (pg. 36)

2.2.6 Molecular Modeling of MosA

Starting with the *E. coli* model, non-identical residues were changed to the correct MosA sequence. Modeling of the MosA structure was carried out using COOT (Emsley, P. and Cowtan, K., 2004) and refinement of the model was carried out using CNS (Brunger et al., 1998). Refinement was carried out from 10.5 Å to 1.95 Å with 20 cycles of conjugate gradient minimization per round of refinement. Final refinement statistics can be found in Table 3.2.

Table 2.3 Table of data processing statistics for MosA crystal data collections. Numbers in parentheses refer to the highest resolution shell.

MosA crystal	Orthorhombic	Trigonal
Wavelength (Å)	1.33 (CLS)	0.90 (APS)
Resolution limit (Å)	19.75 – 1.95 (2.00 – 1.95)	50-3.2 (3.31 – 3.2)
Space group	C 2 2 2_1	P 3_1
No. of molecules/asym unit	2	4
Unit cell dimensions	*a* = 68.9 Å *b* = 138.7 Å *c* = 123.2 Å	a = 115.15 Å c = 95.94 Å
Volume of unit cell ($Å^3$)	1.18 x 10^6	1.10 x 10^6
Specific Volume (mL/g)	1.32	1.24
No. of observed reflections	254154	241480
No. of unique reflections	41799	23381
Data redundancy	6.1	10.3
Completeness (%)	93.4 (74.4)	99.7 (99.9)
I/σ(I)	8.5 (3.9)	3.5 (3.8)
R_{sym}	0.10 (0.26)	0.12 (0.32)

R_{sym} is defined as $R_{sym} = \sum \left| \mathbf{I}-\langle \mathbf{I} \rangle \right| / \sum \mathbf{I}$, where I = individual intensity measured and <I> = average of symmetry-related intensities.

2.3 Comparison of MosA to other NAL Superfamily Members

For comparisons of MosA to other DHDPS structures in the protein data bank, DaliLite was used with homologous regions between MosA and the other DHDPSs being overlaid (Holm and Park, 2000). Root mean square deviation (r.m.s.d.) values were obtained by comparing protein backbone atoms of each subunit to its corresponding subunit in the search model as listed in Table 3.3 (pg. 40). Matrices for transforming pdb coordinates for overlay pictures were generated by DaliLite (Holm and Park, 2000); pdbset, from the CCP4 suite, was used for the transformations (CCP4, 1994). Pictures of the active site and other overlays were generated using Pymol (Delano, W. L., 2002).

2.4 Dynamic Light Scattering

MosA protein in buffer containing 10 mM Tris-HCl pH 8.0, 50 mM KCl, 2 mM 2-mercaptoethanol, and 5 mM pyruvate was filtered through a 0.1 μm Anodisc 13 filter (Whatman) and 20 μl was placed in the testing cuvette of the dynamic light-scattering instrument (DynaPro-MS800). Data were processed using the software supplied with the instrument (*Dynamics* v. 5.26.60, Protein Solutions Inc.). Measurements of the hydrodynamic radius were recorded at intervals from 277 to 318 K.

3.0 Results

3.1 MosA

3.1.1 Protein Purification

After following the procedure for protein purification laid out in Section 2.2.1, Chris Phenix in the Palmer laboratory washed the Ni^{2+} affinity chromatography column with 95% washing buffer (20 mM Tris buffer, pH 8.0, 5 mM imidazole, 500 mM NaCl, and 12.5% v/v glycerol) and 5% elution buffer (washing buffer plus 100 mM EDTA). This purified MosA almost to homogeneity as shown in Figure 3.1. MosA protein was pooled from fractions, dialyzed as described in Section 2.2.1, and frozen as described in Section 2.2.1.

3.1.2 Crystallization and Cryoprotection

Orthorhombic crystals of MosA grew reproducibly overnight as described in Section 2.2.2. Single crystals grew as bipyramids to maximum dimensions of 0.3 x 0.3 x 0.5 mm (Figure 3.2). These crystals were then soaked in cryoprotectant containing lysine and pyruvate as inhibitor and substrate, respectively. Although other cryoconditions were tried, the best conditions were as mentioned in 2.2.3. Crystal soaking conditions that used lower ammonium sulfate concentrations were unsuccessful due to the crystals cracking.

MosA crystals that grew in a trigonal space group using the same crystal growth conditions laid out in 2.2.3. These crystals diffracted to 3 Å but were not pursued further since the orthorhombic crystals diffracted to a higher resolution.

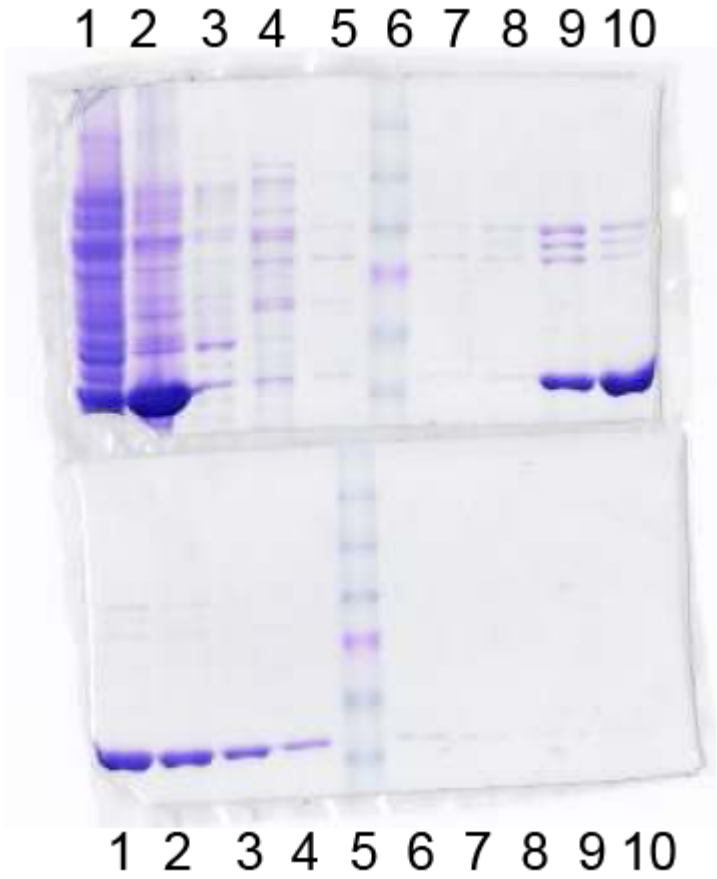


Figure 3.1 SDS-PAGE of MosA purification. **Top gel**: Lane 1, insoluble pellet after sonication; Lane 2, soluble fraction after sonication; Lane 3, flow fraction 1; Lane 4, binding buffer wash; Lane 5, wash buffer fraction 1; Lane 6, protein molecular weight standards (Invitrogen) - molecular weight markers from top of gel as follows 1: 181.8 kDa 2: 115.5 kDa 3: 82.2 kDa 4: 64.2 kDa (pink band) 5: 48.8 kDa 6: 27.1 kDa; Lane 7, wash fraction buffer 2; Lane 8, elution fraction 1; Lane 9, elution fraction 2; Lane 10, elution 3; **Bottom gel**: Lanes 1- 4, elution fractions 4-8; Lane 5, protein molecular weight standards; Lane 6 – 8, elution fractions 8 –10.

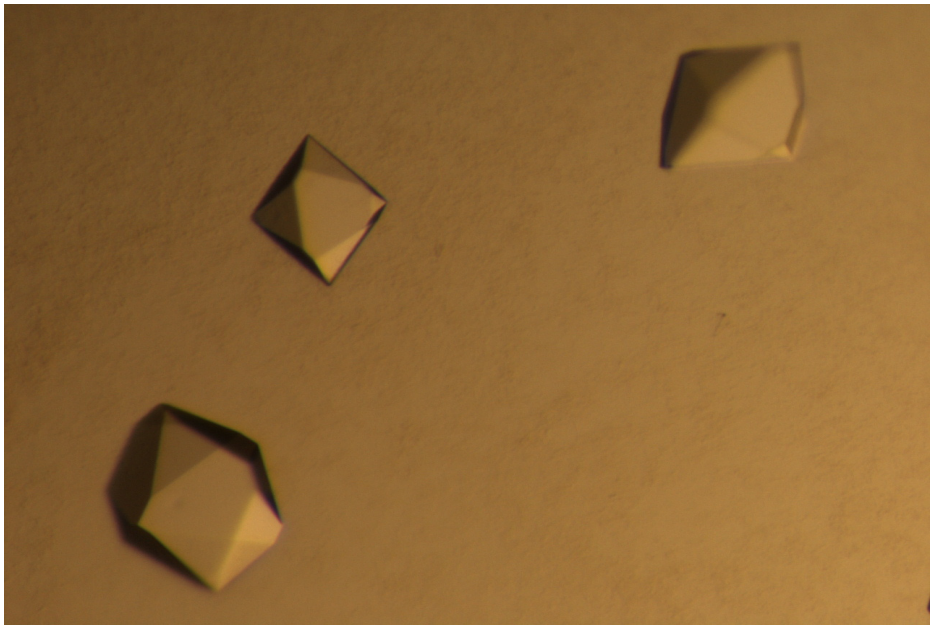


Figure 3.2 Crystals of MosA. Grown overnight in 2 M ammonium sulfate, 100 mM Tris buffer pH = 8.0, and 2% v/v PEG400, the crystals shown are 0.3 x 0.3 x 0.5 mm in size.

3.1.3 Diffraction and Data Processing

Crystals were taken to the Canadian Light Source for diffraction. The first crystal chosen diffracted to 2.2 Å. After problem with the ring current at the CLS, another crystal was chosen for diffraction, as the first data set was incomplete. Following restoration of the x-ray beam, a second crystal was diffracted to a resolution of 1.95 Å. This data set was processed as described in Section 2.2.4 and the structure solved as described in Section 2.2.5.

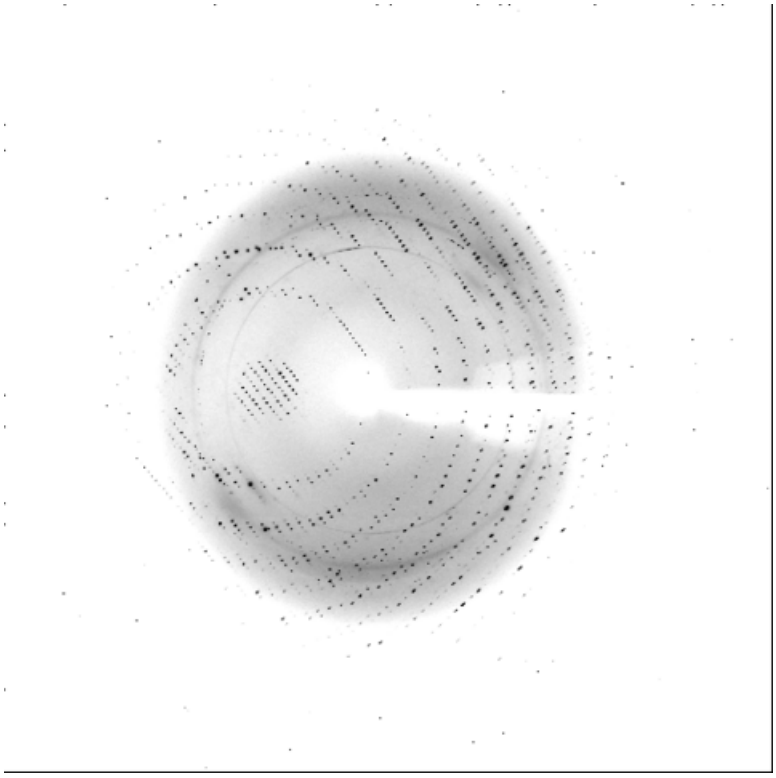


Figure 3.3 Diffraction image from the MosA data collection. Diffraction was observed out to 1.65Å and the data were cut off at 1.95 Å for completeness and merging R factor.

3.1.4 Rotation and Translation Functions

A rotation search was performed using Phaser from the CCP4 suite of programs (McCoy *et al*., 2007). A brute force likelihood method is employed by the program in searching with a starting model as described in Section 2.2.5. A Patterson peak is found corresponding to one MosA protein monomer in the unit cell and is listed in Table 3.1 as 1. Once that solution was selected, brute force translation was used to then find the optimum translation function for the selected rotation. Solutions are picked on the basis of a score assigned to them by Phaser (McCoy *et al*., 2007), and the log-likelihood gain of that solution. The log-likelihood gain is the difference between the likelihood of the model and the likelihood calculated from a Wilson distribution. The likelihood is thus defined as the probability of the data being measured, given the model consisting of the

atomic coordinates and the rotation and/or translation operators applied to those coordinates. In simpler terms, the log-likelihood gain indicates how well the model can predict the data. The log-likelihood gain should be positive and should increase as the solution progresses. A Z-score is also attached to the solution, which is the number of standard deviations over the mean, the mean derived from a score assigned to five hundred random rotations and/or translations. Larger Z-scores indicate more likely solutions, and should increase as the solution progresses. Initial Z-scores are recommended to be over a value of eight.

Once one molecular solution was found, the location of this molecule was fixed. This selected solution was fitted and refined to the final rotation and translation values listed in Table 3.3 as Solution 1. The process of rotating and translating was repeated to find the second molecule, as there are two molecules in the unit cell. The Matthews coefficient was found to be 2.24, giving a solvent content of 45 % with two molecules in the asymmetric unit. Once both rotation and translation solutions were found, they were applied to the starting model and adjusted. Final rotation and translation values are listed in Table 3.3.

3.1.5 Refinement

Once the rotation and translation functions were applied to the starting model, refinement of that model was undertaken as described in Section 2.2.6. The model had an initial R-factor of 41.5%, for both molecules of MosA, and refined successfully from 19.75 – 1.95 Å. The refinement calculations were done using CNS and were carried out after each round of model building (Brunger *et al*., 1998). Final R-factor was 21.9% and

Rfree was 26.8%. The root mean square deviation (r.m.s.d.) for bond lengths and angles was 0.007 Å. All final refinement statistics are shown in Table 3.2 (pg. 38) and the Ramachandran plot (Ramakrishnan and Ramachandran, 1965) shows the strained Y106 conformation (Figure 3.4) (pg. 40).

3.1.6 Structural Overlays

Overlays of MosA and several NAL family members were carried out using lsqkab. DHDPS structures from *H. influenzae, T. maritima*, *M. tuberculosis*, *A. aeolicus* and *E. coli* were superimposed with MosA from *S. meliloti*. All r.m.s.d. values given in Table 3.3 (pg. 41) are for the α carbons overlaid within the given residue range.

Table 3.1 Table of the best rotation and translation search results for MosA search models performed in the space group C 2 2 2_1. Search performed by Phaser from the CCP4 suite (McCoy *et al*., 2007; CCP4, 1994). *Log-likelihood gain (LLG). Initial solutions (top) and final solutions (bottom) for two molecules found in the asymmetric unit. Solution number 3 represents the next highest solution (noise) peak.

Solution number	α	β	γ	transX	transY	transZ	LLG*	Z-score
1	165.2	30.9	97.9	0.908	0.389	0.167	169.95	7.47
2	6.8	78.1	261.9	0.847	0.322	0.889	156.28	7.69
3	2.1	62.2	53.3	0.637	0.682	0.553	15.09	4.39

Solution number	α	β	γ	transX	transY	transZ	LLG*	Z-score
1	167.8	31.2	94.7	0.928	0.412	0.170	191.60	15.41
2	7.1	77.7	262.9	0.838	0.307	0.889	169.65	15.01

Table 3.2 Final refinement statistics for MosA using CNS (Brunger *et al*., 1998). Numbers in parentheses refer to highest resolution shell.

Resolution limits (Å)	10.5 – 1.95 (2.00 – 1.95)
R_{work} (37312 reflections)	0.219 (0.33)
R_{free} (2073 reflections)	0.268 (0.37)
R_{cryst} (39385 reflections)	0.228
No. of residues	584
No. of non-hydrogen protein atoms	4408
No. of water molecules	785
Mean B factor main chain atoms ($Å^2$)	20.6
Mean B factor side chain atoms	22.9
Mean B factor water molecules	34.1
Mean B factor SO_4^{2-} ions	31.1
r.m.s.d. from ideal geometry	
Bond distances (Å)	0.007
Bond angles (˚)	1.49
Dihedral angles (˚)	21.8
Improper angles (˚)	0.91
Ramachandran plot regions	
Residues in most favoured regions (%)	91.2
Residues in favoured (%)	8.0
Residues in additional allowed regions (%)	0.4
Residues in disallowed regions (%)	0.4

$R_{sym} = \Sigma | \langle I_{hkl} \rangle - I_{hkl} | / | I_{hkl} |$, where $\langle I_{hkl} \rangle$ is the average intensity over symmetry-related reflections and I_{hkl} is the observed intensity.
$R_{value} = \Sigma | | F_o | - | F_c | | / \Sigma | F_o |$, where F_o and F_c are the observed and calculated structure factors. For R_{free} the sum is done on the test set reflections (5% of total reflections), for R_{work} on the remaining reflections, and for R_{cryst} on all reflections included in the resolution range.

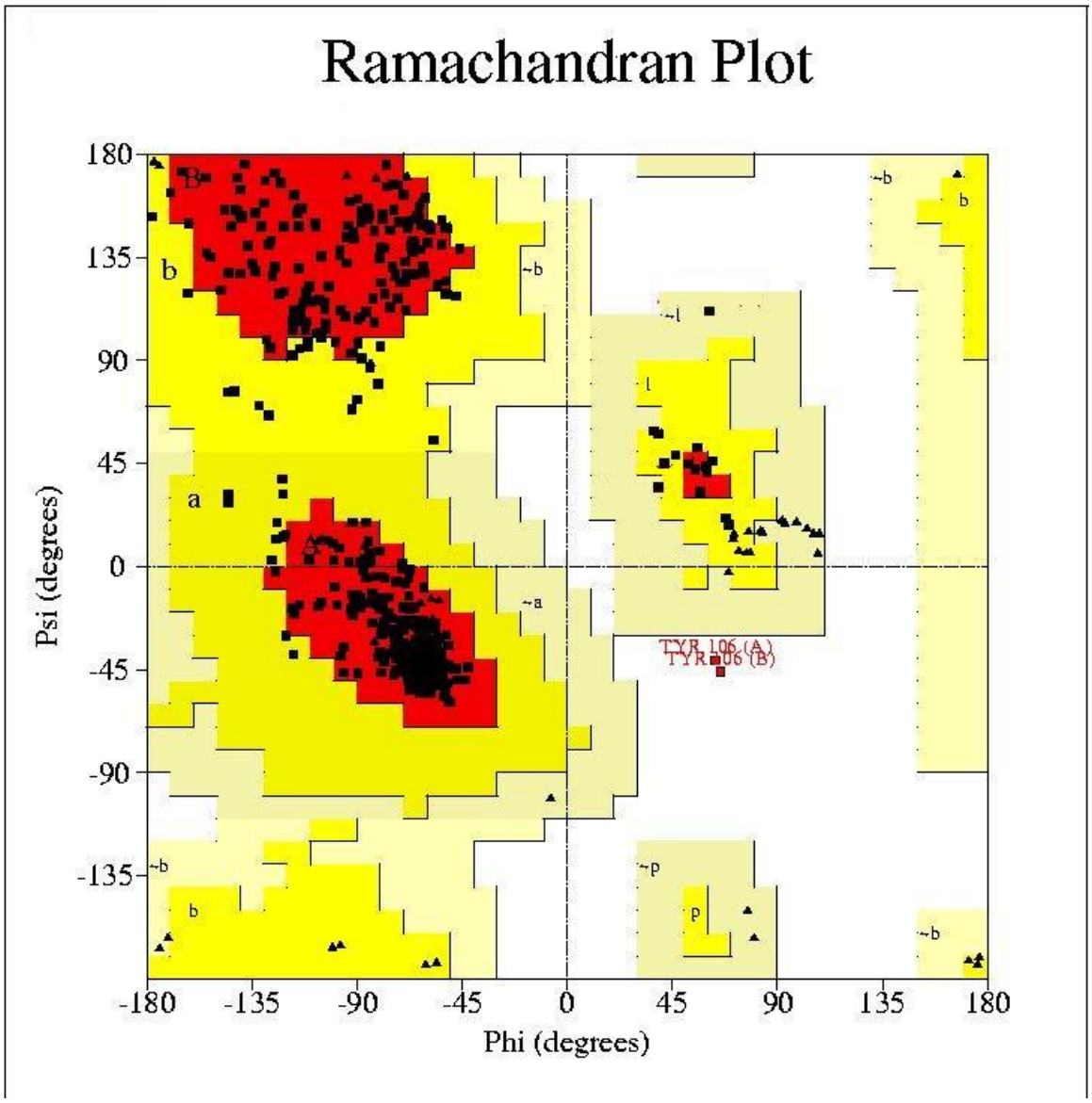


Figure 3.4 Ramachandran plot for the structure of MosA complexed with pyruvate. The red coloured region is the most favoured region, the yellow and light yellow represent favoured and additional favoured regions. Shown in disallowed space are 2 residues, Y106 A and B. Y106 has adopted a well documented strained conformation as seen in other DHDPS structures.

Table 3.3 Final backbone atom structural overlay values for MosA from *S. meliloti* and DHDPSs found in the Protein Data Bank. All r.m.s.d. values generated from DaliLite (Holm and Park, 2000).

PDB code	Organism	Residue Range	% Identity	r.m.s.d. (Å)
1DHP	*E. coli*	54-68; 74-124; 159-175; 202-245; 251-259; 275-288	45	1.2
1YXD	*E. coli* w/ lysine bound	54-68; 74-124; 159-175; 202-245; 251-259; 275-288	45	1.2
1F5Z	*H. influenzae*	58-72; 78-128; 163-179; 205-249; 254-262; 279-292	26	2.3
1F73	*H. influenza* w/ sialic acid	58-72; 78-128; 163-179; 205-249; 254-262; 279-292	26	2.3
1O5K	*T. maritima*	65-79; 85-145; 170-186; 213-256; 262-270; 286-299	40	1.4
1XXX	*M. tuberculosis*	67-81; 87-147; 172-188; 215-257; 264-272; 288-301	34	1.6
2EHH	*A. aeolicus*	53-67; 73-123; 158-174; 201-244; 250-258; 274-287	43	1.3

3.1.7 Dynamic Light Scattering

Dynamic light scattering performed on MosA indicated a highly homogenous solution that was likely to be composed of tetramers, as are other DHDPSs. The data revealed a bimodal distribution with one highly monodisperse (<15% polydispersivity) peak comprising 99% of the mass and having an approximate hydrodynamic radius of 4.2 nm, corresponding to a 130 kDa tetramer. The protein solution tolerated a change of temperature from 277 to 303 K. Above 318 K, the detector recorded an overload indicating a temperature limit had been reached, which was likely to coincide with denaturation of the protein.

4.0 Discussion

4.1 MosA crystal structure

4.1.1 Overall folding

The MosA protein is a tetramer in solution and in the crystal structure. Four subunits come together to form the tetramer in alternating $(\beta/\alpha)_8$ packing in the crystal (Figure 4.1). Consisting of eight parallel β-strands surrounded by eight α-helices, the TIM barrel was first described in triosephosphate isomerase (Banner *et al*., 1975; reviewed by Reardon and Farber, 1995). Each monomer of MosA consists of a TIM barrel with two additional α-helices. The tetramer then provides four active sites as well as a channel between dimers for substrate flow.

4.1.2 Crystal Packing

In the space group C 2 2 2_1, the 2_1 screw axis is, by convention, along the *c*-axis. A 2- fold symmetry is observed along the *a* and *b* axes (Figure 4.2). Due to C-centering, additional 2_1 screw axes are observed at 0, 0, ¼; ¼, ¼, 0; ¼, 0, ¼; 0, ¼, ¼. There are two molecules in the asymmetric unit and thus sixteen monomers in the unit cell.

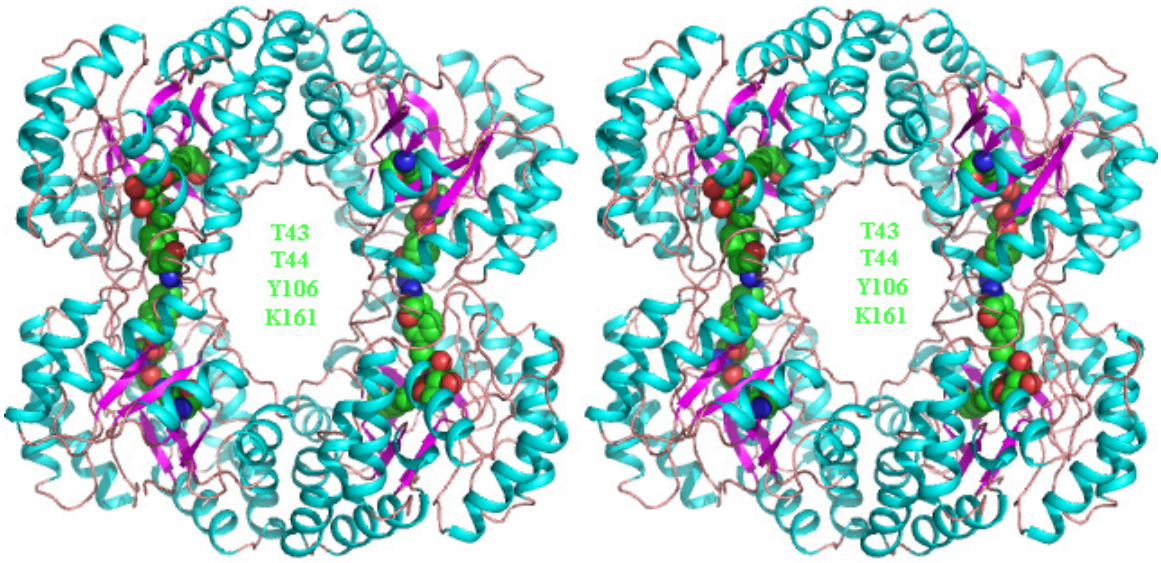
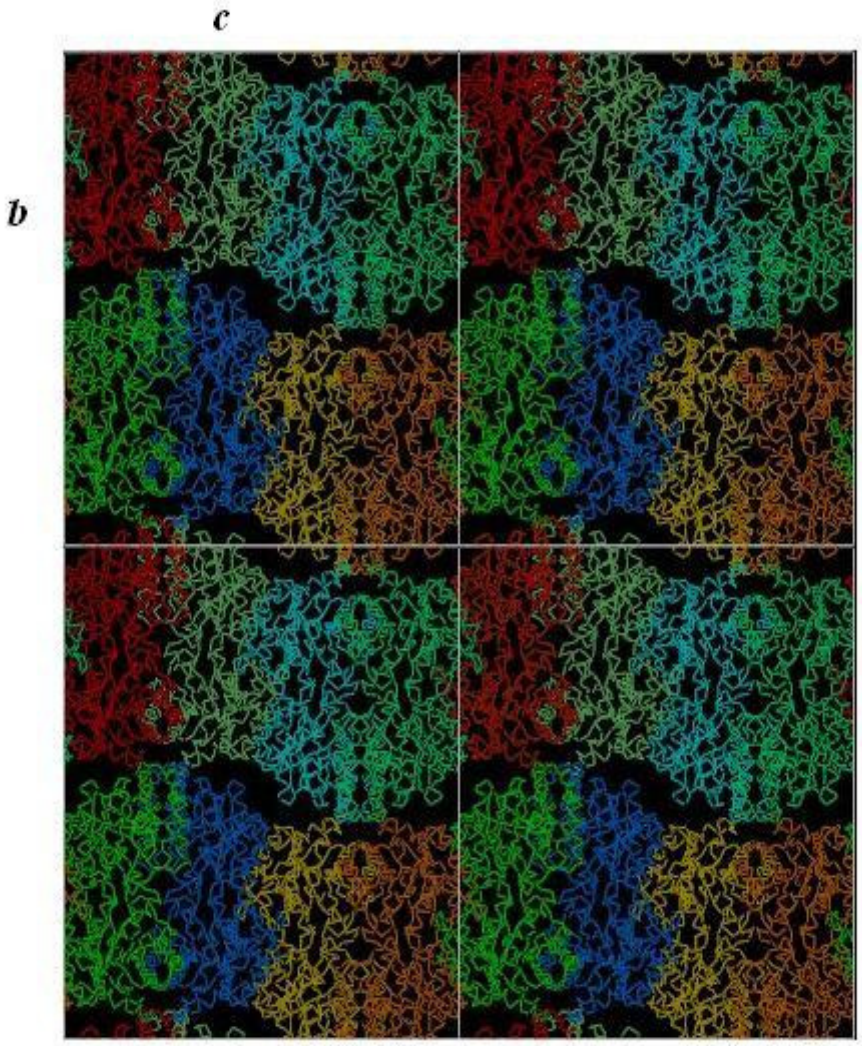
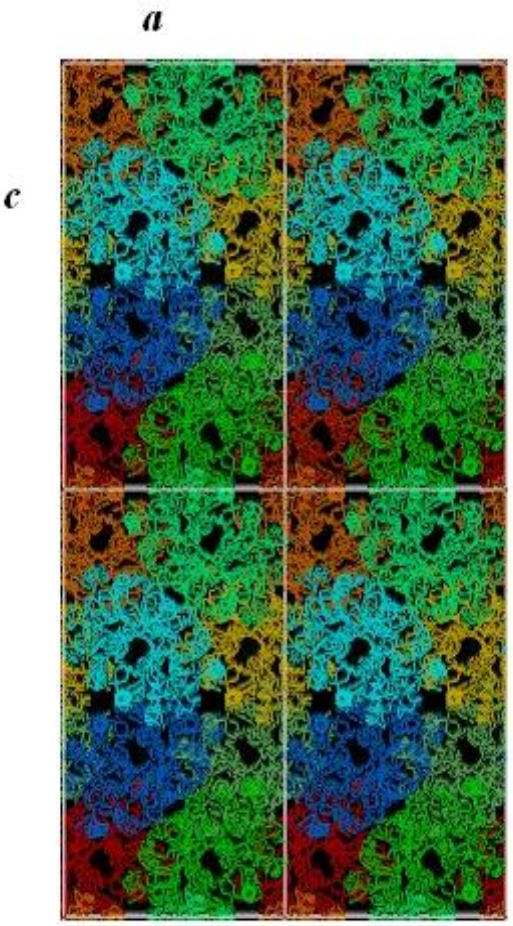


Figure 4.1 Stereo view of the MosA tetramer. Visible is the overall TIM barrel folding and active site orientation. Y106 reaches from one subunit into the adjacent one, linking the two subunits as seen above. Pictured in spheres, key active site residue carbons are coloured in green, oxygens red, and nitrogens blue.


c
b


a
c

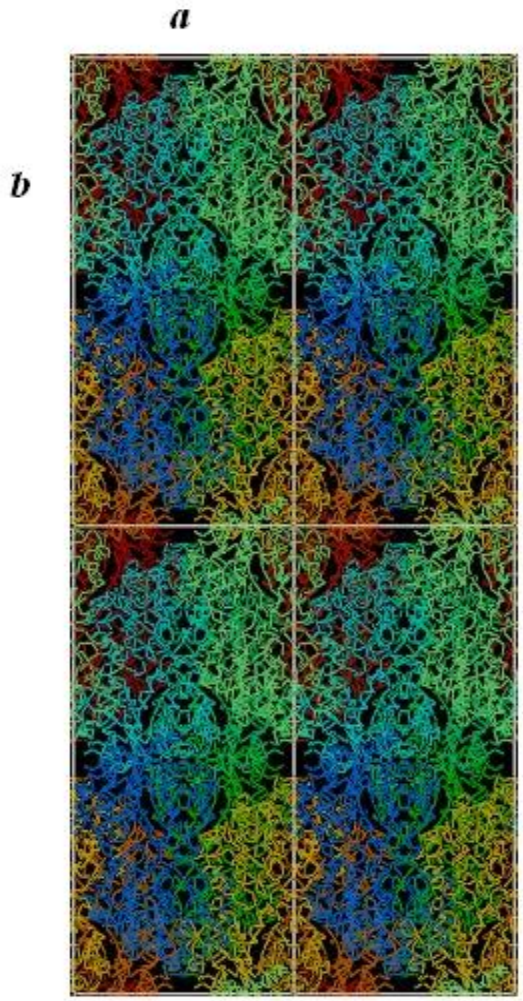


Figure 4.2 Crystal packing of MosA. Depicted are unit cells as viewed down the *a, b,* and *c* axes of the unit cell. Image made using TURBO-FRODO (Roussel, A. and Cambillau, C., 1990).

4.1.3 Active Site Geometry

The MosA active site is centered on residue K161 (Figure 4.3). In the first step of the lyase reaction, pyruvate is covalently attached to K161 in a dehydration step, resulting in an imine formation. Y106, Y132, and T43 form a catalytic triad that is highly conserved in other DHDPSs. Other DHDPS structures have reported the involvement of R137 in substrate binding. In the structure reported here, R137 is located over 5 Å away from the nearest active site residue, Y132. It is hypothesized this arginine is acting as a gatekeeper to the active site hole that is located in the centre of the β-barrel (Dobson *et al*., 2005b). *L*-ASA binds after pyruvate, and a possible role for R137 centres on this second substrate binding (Figure 4.3, Figure 4.4).

K161, labelled to as LYP in the structure reported here with pyruvate bound in imine form, forms two hydrogen bonds with the side chain oxygens of T43 and T44. A list of all hydrogen bonds in the active site can be found in Table 4.1. There are another two hydrogen bonds formed with the peptide nitrogens of the same residues. In addition, T43 is highly conserved in the NAL superfamily. This is unsurprising as hydrogen bonds are formed between it and the two tyrosines, Y106 and Y132, two of the key residues in the catalytic triad forming a proton relay. It is also worth noting that Y106, stretching over from the other subunit in the dimer, shows a strained carbonyl oxygen conformation. This means the hydroxyl group is oriented toward T43 forming a hydrogen bond and positioning the tyrosine for proton transfer. In addition to that, a hydrogen bond is formed between the strained carbonyl oxygen of Y106 and the epsilon nitrogen in R137. Proper positioning of the arginine for binding to *L*-ASA may thus be accomplished through the strained conformation adopted by Y106. As well, proper positioning of

Y106 may be accomplished by R137, although the variability of R137's position in previously solved structures and the inflexibility of Y106 lends credence to the role of Y106 in R137 alignment and not necessarily the other way around.

4.1.4 Description of pyruvate Schiff base adduct

Covalently bound to the key active site residue K161 is pyruvate that has been converted to the imine form. Located in close proximity to the catalytic triad, K161 is centered at the bottom of the hole created by the β-barrel (Figure 4.4). This is the first time a pyruvate molecule has been reported in the active site of a DHDPS (Figure 4.5). The conversion to imine form lends credence to the previously postulated enzyme mechanism (Figure 1.8) (pg. 12).

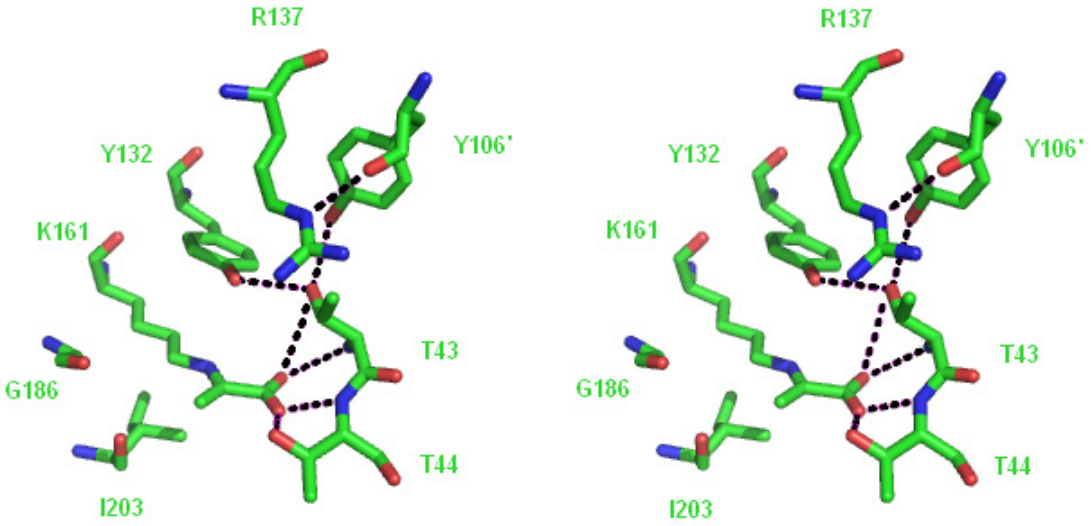


Figure 4.3 Stereo view of the MosA active site. Covalently bound to K161 following conversion to an imine is a pyruvate molecule. R137 is pictured in the upper portion near Y132 at the opening of the active site. H-bonds are denoted as black dashes.

Table 4.1 Table of hydrogen bonds in the MosA active site.

Residue	Residue	Distance (Å)
K161 O1	T44 OG	2.69
K161 O1	T44 N	2.89
K161 O2	T43 OG	3.13
K161 O2	T43 N	3.05
Y132 OH	T43 OG	2.83
Y106’ OH	T43 OG	3.04
Y106’ O	R137 NE	2.45

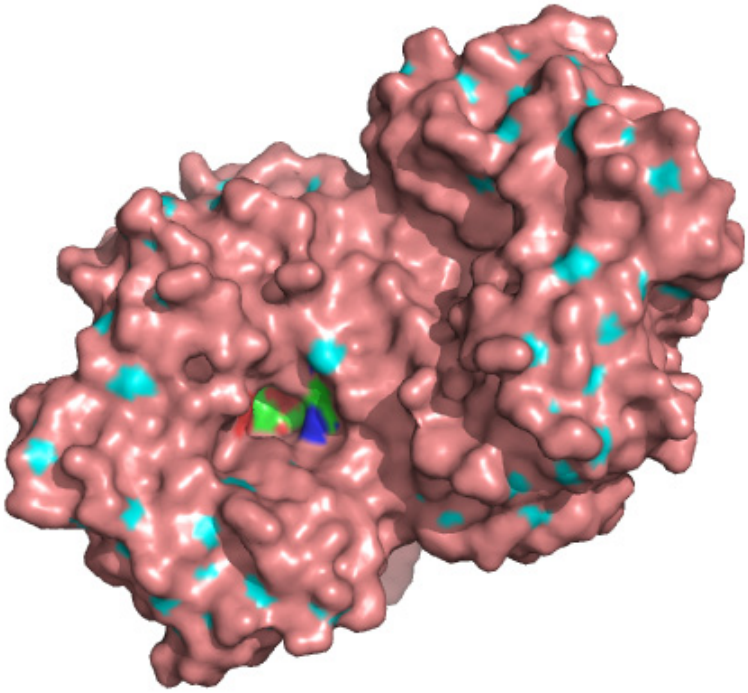


Figure 4.4 Surface view of the MosA dimer. The active site is coloured in the centre by atom type, carbon being green, oxygen is red, and nitrogen blue. The surface is coloured by secondary structure, with β-strands in pink and α-helices in light blue. The active site is visible in the centre of the β-barrel hole. The side chain of R137 is just visible on the right side of the hole, and in other MosA structures (not reported) is present over the active site indicating a possible gate-keeper role.

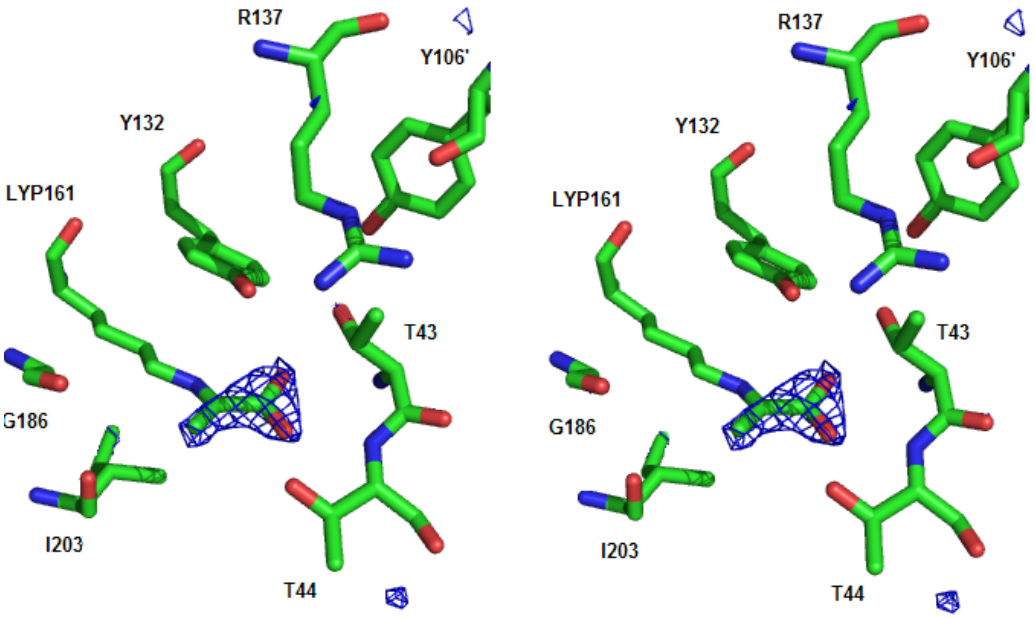


Figure 4.5 Pyruvate bound at the MosA active site. An omit map was calculated, following refinement, with the pyruvate adduct removed from the structure factor calculation and is shown here contoured at 3 σ.

4.1.5 New Proposals for Enzyme Mechanism

Previous structures have reported waters or chloride ions bound in the active site. As this structure has pyruvate covalently bound, a closer examination of the enzyme mechanism is possible. The currently accepted mechanism involves formation of an imine of pyruvate with K161, which as discussed, is seen in this MosA structure. In the proposed mechanism, Y132 acts as an acid and final stop in the proton relay between solvent and pyruvate. Y132 is also involved in the subsequent reaction with *L*-ASA, acting again as an acid donating a proton, later regaining its proton acting as a base in a further reaction with the enzymatic product. G186 is also involved in another important function, anchoring the *L*-ASA to the active site. Eventually, G186 aids in product formation by orienting the para-hydroxyl group in the axial position of the 4-hydroxytetrahydrodipicolinic acid.

Binding of *L*-ASA may be facilitated by movement of R137 but without a crystal structure showing movement, this would be pure speculation. It is believed that given the location of R137, and investigation of entropies by Dobson *et al*., the side chain of R137 is highly likely to be involved (Dobson *et al*., 2005a). Overlays between the *E. coli* structure and MosA do not show any significant difference in side chain location when pyruvate is bound, indicating a role in substrate movement in and out of the active site, or as mentioned above, a role in *L*-ASA binding. Investigation with mutants of this amino acid would shed light on this debate.

4.2 MosA and other DHDPSs

4.2.1 *E. coli* DHDPS

Given the 45% amino acid sequence identity between *E. coli* DHDPS and *S. meliloti* MosA, the fact that their backbone atoms align to 1.2 Å is of little surprise. Overall, the Cα trace can be superimposed as shown in Figure 4.6. Even when superimposing MosA and *E. coli* DHDPS with added free lysine bound, the structure shows little difference, with an r.m.s.d. value of 1.2 Å (Table 3.3)(pg. 40) (Figure 4.6, 4.7, and 4.8).

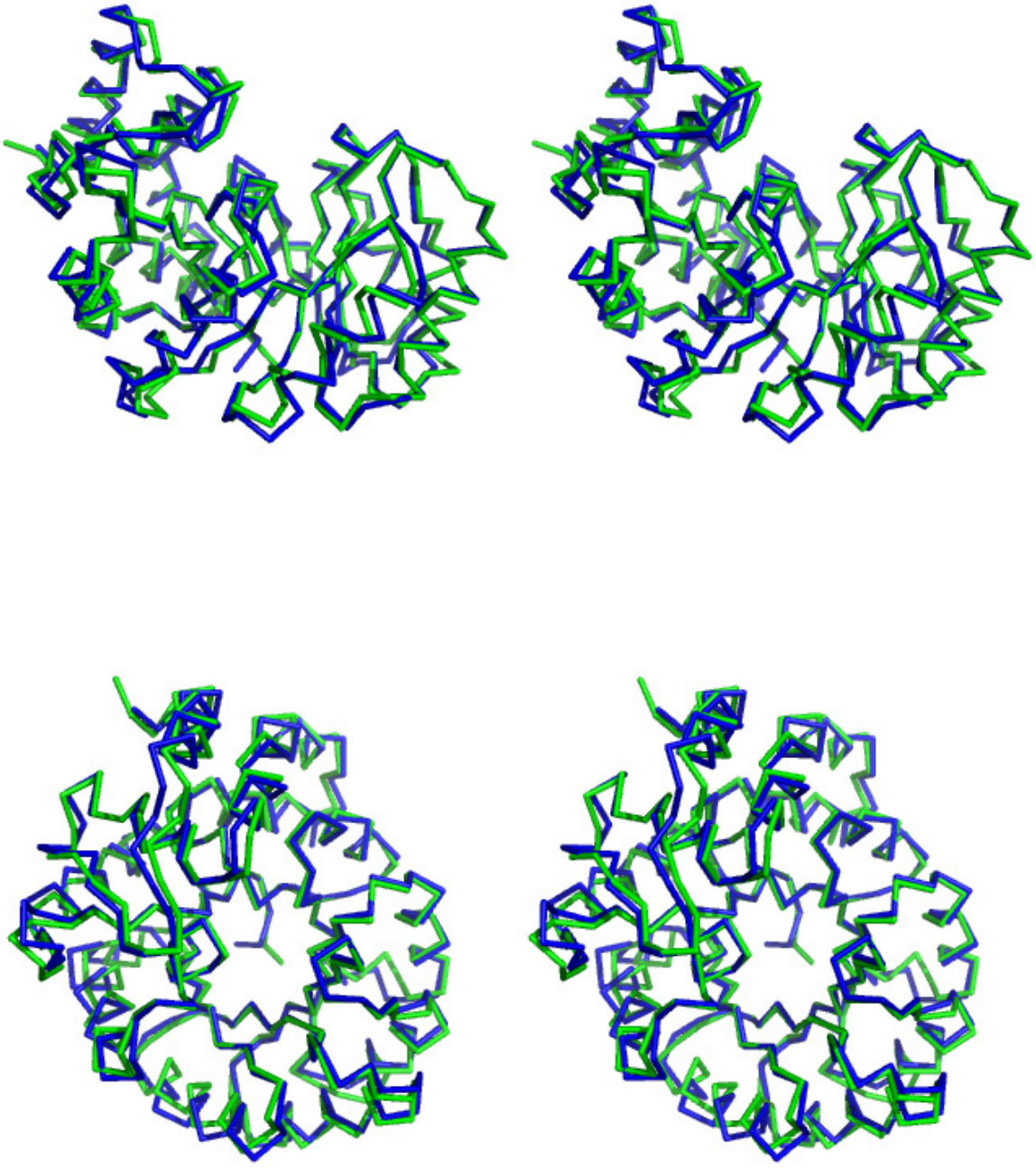


Figure 4.6 Stereo diagram of the overlays of *E. coli* DHDPS (1DHP) and MosA backbone atoms. MosA is depicted in green, *E. coli* DHDPS in blue.

4.2.1.1 Inhibition by lysine

The *E. coli* DHDPS structure reported by Dobson *et al* has two single lysine amino acids bound at allosteric sites (Figure 1.9 and 4.8) (Dobson *et al*., 2005b). This is supported by work in the Palmer lab that shows binding of 2 lysines for every one pyruvate (Tam *et al*., 2004). Binding of an inhibitor at an allosteric site did not confer any significant structural change to the active site. Allosteric inhibitors, by definition, bind at a site in the enzyme away from the active site and lower the affinity of the enzyme for the substrates. Previous investigation indicated that lysine might inhibit DHDPS by inhibiting R138 (*E. coli* numbering) movement (Blickling *et al*., 1997). In the Dobson structure, this was found not to occur.

A second theory has been put forward regarding lysine inhibition. In examining the structure of DHDPS from *E. coli*, researchers concluded that lysine inhibition was facilitated by blockage of a solvent channel, thereby trapping reaction end-products in the active site (Dobson *et al*., 2005b). In the structure reported here, no such solvent channel can be found. In addition, when overlaid with the Dobson structure, MosA with pyruvate bound shows neither significant side chain location difference, nor blockage of any channel (Table 3.3) (Figure 4.8). This lack of movement is in agreement with work done by Chris Phenix in the Palmer laboratory (Phenix, 2007). His investigation of MosA using isothermal titration calorimetry (ITC) revealed that lysine binding to MosA when pyruvate was bound was enthalpy driven. This finding indicates that when pyruvate is present, enthalpic heat is being released that drives binding, not the exclusion of water molecules. Indeed, when pyruvate is not bound to MosA, lysine binding is entropically driven, indicating water molecules in and around the allosteric binding site are released.

The end result of the ITC work is that MosA binds lysine when pyruvate is bound because it is enthalpically favourable, not because it is entropically favourable. The expulsion of waters from the active site or blockage of a solvent channel does not appear to be the mechanism of lysine inhibition. What can be said is that lysine is a mixed inhibitor of MosA, binding to both E and ES complexes. This supports kinetic work done by Dobson and contraindicates other work that shows lysine is an uncompetitive inhibitor with respect to pyruvate (Dobson *et al*., 2005b). Overlaying the *E. coli* structure, with lysine bound, to this structure of MosA with pyruvate bound has not produced any clearer results, despite an r.m.s.d. of only 1.2 Å. Without a structure of lysine and pyruvate bound to MosA, the mechanism of inhibition remains unresolved.

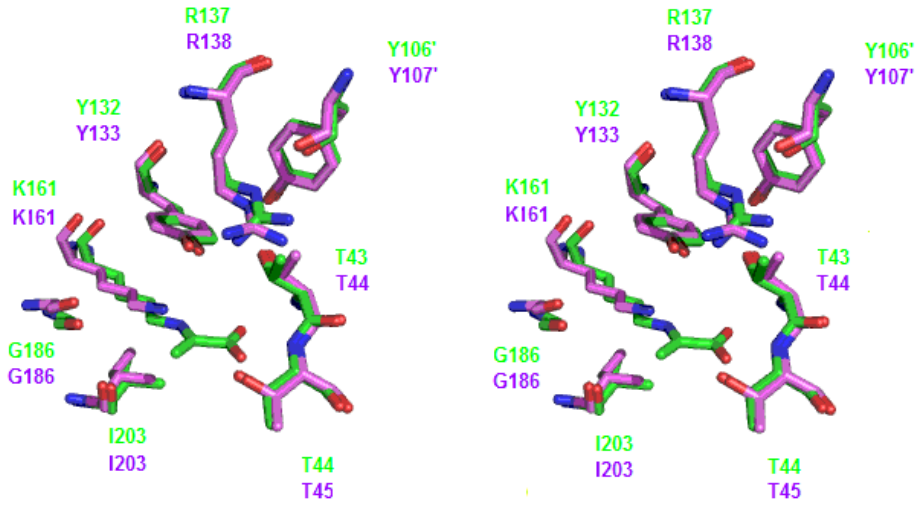


Figure 4.7 MosA active site with *E. coli* DHDPS active site, with lysine bound (not pictured), overlaid. Overall r.m.s.d. was 1.2 Å indicating a close alignment of alpha carbon atoms. As pictured, overall side chain movement is essentially negligible. MosA is shown in green, *E. coli* DHDPS in purple.

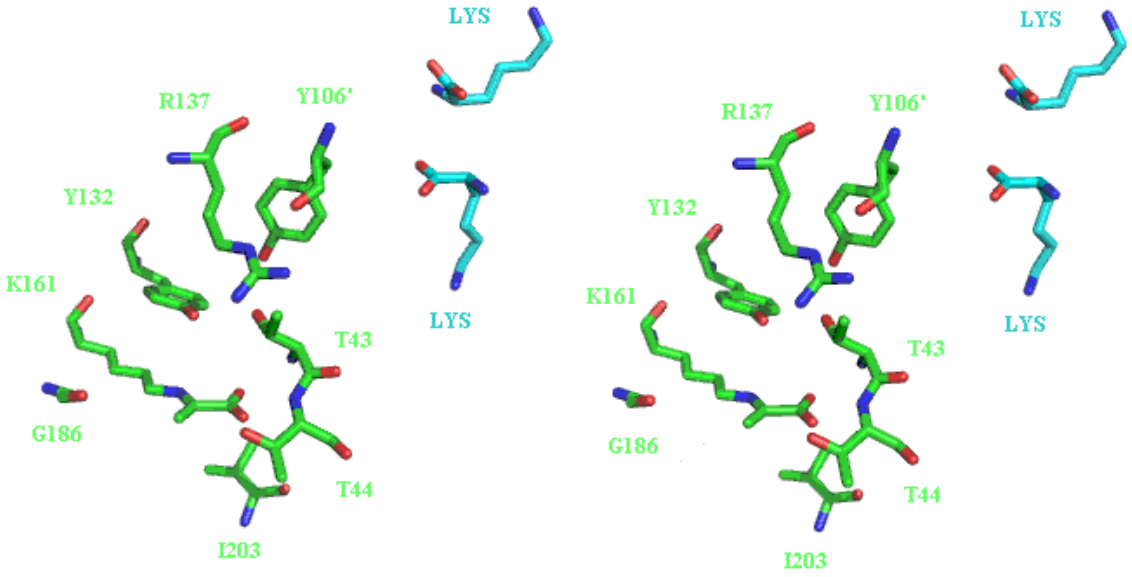


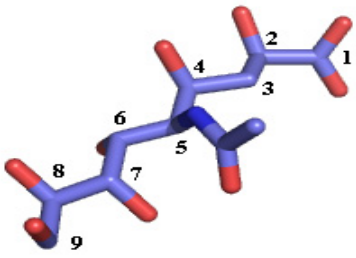
Figure 4.8 MosA active site with relative location of the two solitary lysine amino acids from *E. coli* DHDPS superimposed. Visible on the right, the two individual lysines are located in the allosteric binding site. MosA is depicted in green, the lysines from *E. coli* DHDPS complex in blue.

4.2.2 *Haemophilus influenzae* N-acetylneuraminate lyase

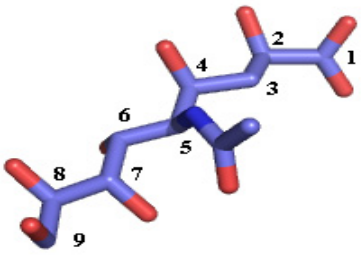
MosA is a member of the NAL superfamily so while its reaction may differ from other family members; its structural framework is similar. As such, structural alignments were carried out with a 2.3 Å r.m.s.d. between MosA and *H. influenzae* DHDPS. Similar to the *E. coli* overlay, the Cα trace is very similar. Overall folding remains the same with tetramers present and similar contacts made between interfaces.

4.2.2.1 Inhibitor bound complexes

Lawrence *et al*., in 2000, reported structures with 3 substrate analogs, including sialic acid alditol (2,4,6,7,8,9-hexahydroxy-5-methylcarboxamido nonanoic acid : HMN) bound in the active site of *H. influenzae* NAL and not at the allosteric site (Figure 4.10). The carboxylic acid group of the inhibitor aligns in the same location and orientation of the pyruvate carboxylic acid group bound in the MosA active site. The inhibitor HMN, positioned in the MosA active site, makes hydrogen bonds with a conserved threonine residue seen in the MosA structure with pyruvate (Figure 4.10).



HMN



HMN

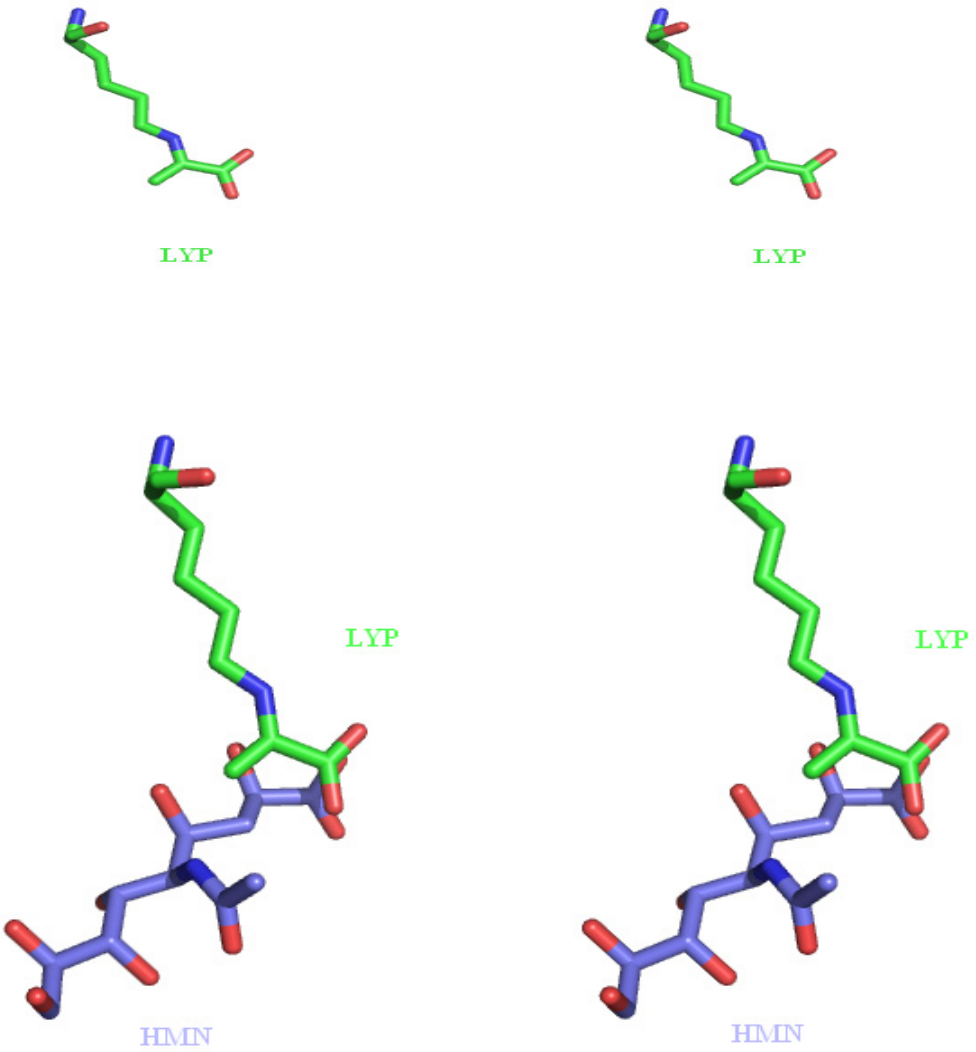


Figure 4.9 After overlapping backbone atoms of *H. influenzae* NAL with MosA, the stereo view of HMN (sialic acid) bound to the *H. influenzae* NAL, and the Schiff base adduct formed by lysine and pyruvate. The HMN acts as an inhibitor in *H. influenzae* and is shown overlaid with K161 of MosA. Of note, the two oxygens of HMN which overlap the pyruvate and lysine adduct oxygens in MosA, thereby making key hydrogen bonds and removing the possibility of pyruvate binding to the *H. influenzae* NAL active site.

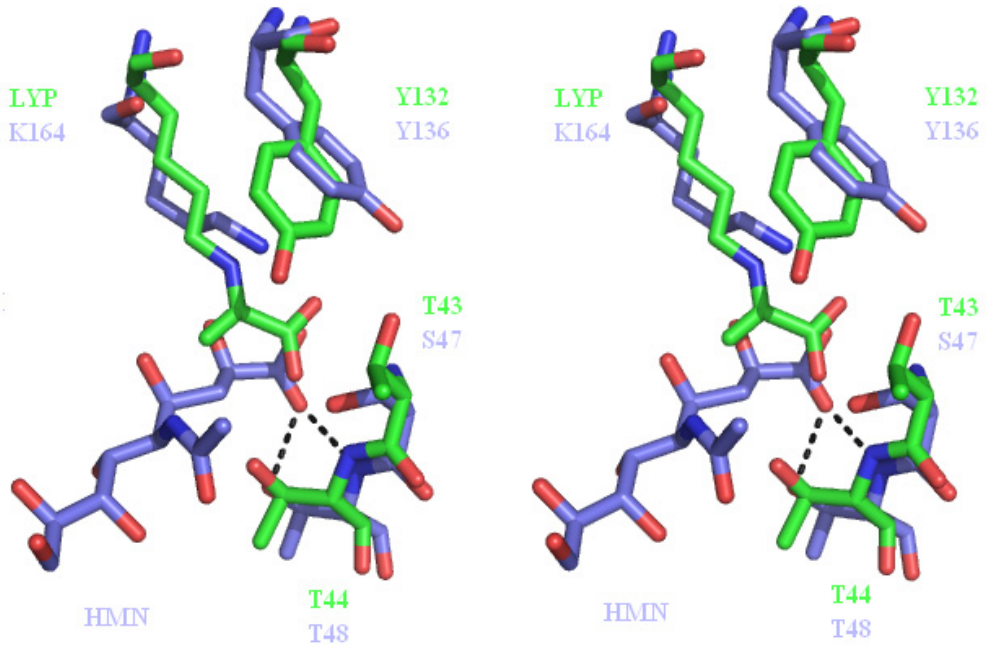


Figure 4.10 After overlapping the backbone atoms of the protein structures, a stereo view of key active site residues from MosA and NAL from *H. influenzae*. HMN, or sialic acid alditol, is also shown in its bound position, making key hydrogen bonds, shown in black. MosA is pictured in green, H. influenzae NAL in dark blue.

As mentioned, the inhibitors bound to the *H. influenzae* NAL structure bind directly in the active site, and in two cases directly to the active site lysine via covalent bond, so these overlapping atom sites are to be expected. Of note in the sialic acid alditol (HMN) bound structure is the movement of key active site residues. As shown in Figure 4.10, Y136 is shown to be moved away from the inhibitor. The mechanism of this inhibitor is two-fold. First, simply by occupying the pyruvate binding site, and making key hydrogen bonds, HMN, in *H. influenzae* NAL, is able to inhibit the NAL reaction before the first step in the reaction. The second method of action seems to involve displacement of key residues, in this case the active site lysine and tyrosine side chains, and would disrupt the stacking and proton transfer ability of the catalytic triad in the enzyme. This does confirm the ability of NAL superfamily members to be inhibited and lends further credence to the currently accepted mechanism of action for MosA.

4.2.3 Other DHDPSs comparison to literature

The most studied DHDPS in the literature is from *E. coli*. As mentioned in section 3.1.6, structural overlays between *A. aeolicus, M. tuberculosis, T. maritima,* and *H. influenzae* were performed (Figure 1.5). No significant changes or movements can be observed in the active site and overall folding shows the same TIM barrel motif and tetramer formation observed in MosA. Interestingly, the one member of the family showing markedly different tetrameric arrangement has only one example. DHDPS from *N. sylvestris* adopts the reverse quaternary formation from all other known DHDPSs (Blickling *et al*., 1997) (Figure 4.11).

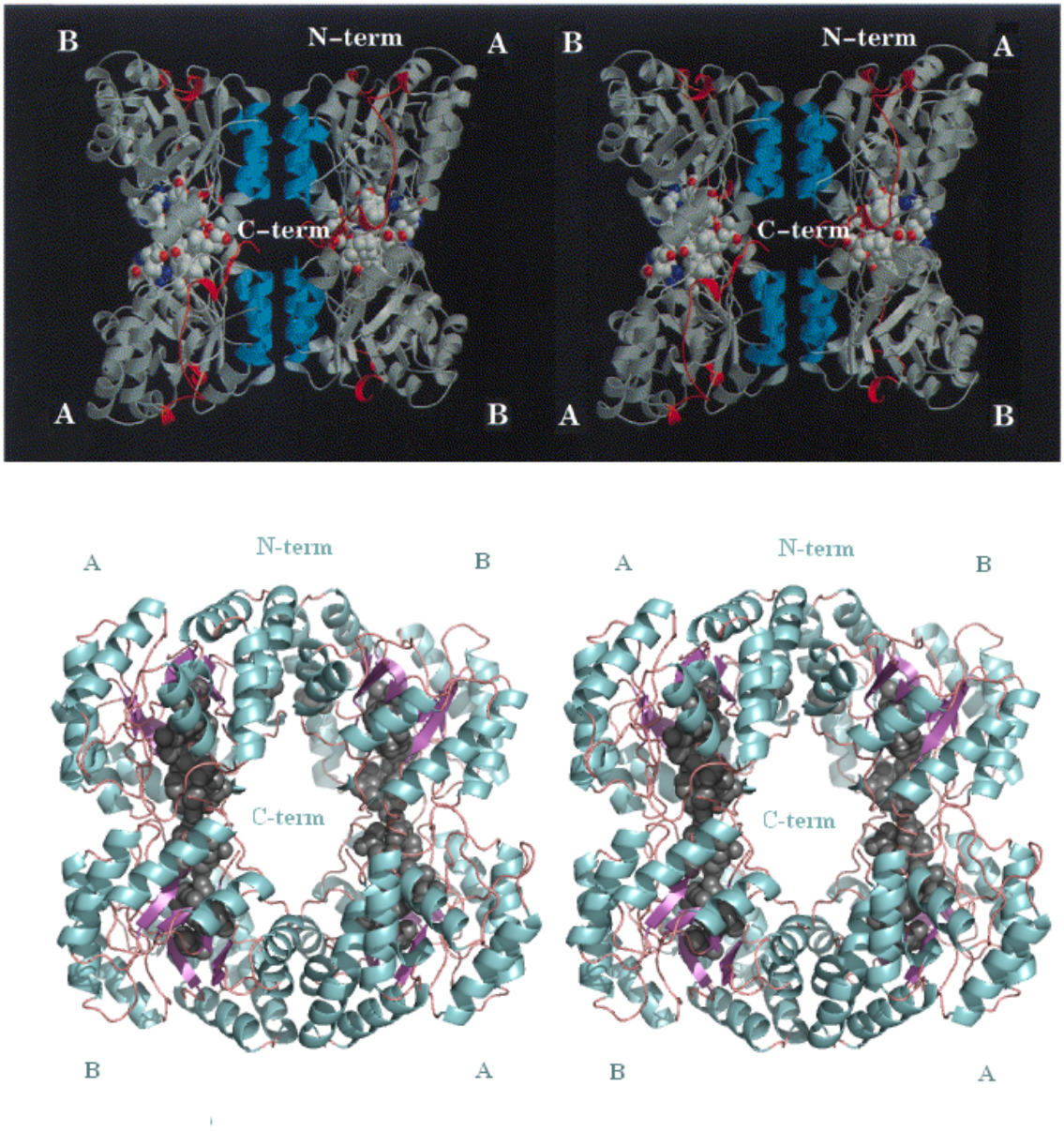


Figure 4.11 Stereo view of the *N. sylvestris* DHDPS (top) (adapted from Blickling *et al*., 1997) and MosA (bottom). Active site residues are illustrated as spheres in *N. sylvestris* DHDPS and MosA. The tetramer formation of *N. sylvestris* DHDPS is different from all other known DHDPSs. As shown, MosA adopts the conformation observed in all other solved DHDPS structures. Lettering denotes chain identifier.

As seen in Figure 4.11, the dimer interface between subunits remains unchanged, but the tetramer interface is on the opposite side to as compared to MosA and to other DHDPSs. This orients the active sites of *N. sylvestris* DHDPS out into the cytosol and significantly narrows the solvent channel between the dimer pairs. The reasoning for adopting this conformation remains unknown, and catalytic efficiency remains unchanged.

5.0 Conclusions and Future Studies

5.1 MosA from *S. meliloti*

5.1.1 Enzymatic characterization

In this thesis, the crystal structure of MosA from *S. meliloti* is reported. This enzyme was originally characterized as a methyltransferase (Rao *et al*., 1995). A sequencing error was discovered and corrected, placing MosA in the NAL superfamily of enzymes as a DHDPS (Tam *et al*., 2004). The structure reported here shows pyruvate, the first reactant in the DHDPS reaction, bound covalently to the active site. There is no longer a question of MosA functioning as a methyltransferase. The main reaction catalyzed is the DHDPS reaction, as confirmed by this crystal structure and work done in the Palmer lab using ITC.

The x-ray structure reported here also helped confirm the proposed mechanism of action for the DHDPS reaction. The pyruvate is captured in the active site, forming an imine with K161, as predicted in the currently accepted mechanism (Figure 1.8) (pg. 12). In addition, hydrogen bonding and close proximity of conserved active site residues confirms their role in the catalytic triad and in the proposed proton relay in the mechanism.

5.1.2 MosA and other NAL superfamily members

In examining the similarities of MosA to other related enzymes, structural overlays were performed to shed light on key structural features. All of the NAL superfamily enzymes that have had their Cα atoms overlaid within 2.3 Å r.m.s.d. This similarity indicates the importance of the overall tetrameric structure as previously reported. The TIM barrel motif allows solvent to flow into, and away from, the active site, and the "reaching across" of Y106 from one subunit to another, allows a tight dimer interface for proper orientation and enzymatic activity.

5.1.3 MosA inhibition by lysine

Although attempts were made, the soaking of lysine in MosA crystals did not produce a visible adduct. However, given the degree of similarity and overlay of 1.2 Å r.m.s.d., comparing the position of lysine in the *E. coli* DHDPS structure with MosA, reasonable theories can be formulated as to the mechanism of inhibition. Also given that MosA is similar to other NAL family members, any inhibitor of MosA should likely be applicable in other organisms that are classified as pathogens, and thus MosA becomes a good structure for study in drug and herbicide development. Although the exact mechanism is unknown, the structural data combined with ITC data make a case for disruption of aromatic stacking in the active site. As seen in the NAL structure from *H. influenzae*, inhibitors acting in such a fashion can be successful. Although lysine is an allosteric inhibitor, subtle changes in active site geometry may ultimately inhibit the DHDPS reaction.

5.2 Future work with MosA

The mechanism of lysine inhibition of MosA remains a question. Any future work in this area should revolve around a structure containing both pyruvate and lysine. As mentioned, crystal soaking in this case was unsuccessful, so co-crystallization remains as the best option for introduction of these ligands. In addition, a structure with *L*-ASA bound may provide further insight or corroboration for the currently accepted enzyme mechanism. There are questions remaining about the role of R137, specifically because it can be observed in a variety of locations in the active sites of the NAL family members whose structures have been solved. In looking at a surface model of MosA, R137 is positioned over the active site cavity, making it ideal to bind *L*-ASA or cover the active site once *L*-ASA is bound. Arginine has a flexible enough side chain to allow movement before and after product formation, making this amino acid another target for mutation studies, and making it a possible inhibitor target as well.

Future work with ITC may also provide enzymatic characterization in the absence of further crystal structures. A mutant MosA with changes to known or suspected active site residues would have interesting binding affinities and enzyme kinetics. ITC has proven to be a valuable tool for such examinations. Work in the Palmer laboratory has already shown ITC can rule out interactions of potential substrates, and has proven enthalpically versus entropic driven interactions can provide insight to mechanism, even with a known crystal structure.

6.0 References

Babbitt, P. C. and Gerlt, J. A. (1997). Understanding Enzyme Superfamilies. Journal of Biological Chemistry, *272*, 20591-30594.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975). Structure of Chicken Muscle Triose Phosphate Isomerase Determined Crystallographically at 2.5 Å Resolution Using Amino Acid Sequence Data. Nature, *255*, 609-614.

Barbosa, J. A. R. G., Smith, B. J., DeGori, R., Ooi, H.C., Marcuccio, S. M., Campi, E. M. *et al.*, (2000). Active-site modulation in the *N*-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed *Haemophilus influenzae* enzyme. Journal of Molecular Biology, *303*, 405–421.

Blickling, S., Renner, C., Laber, B., Pohlenz, H., Holak, T. A., and Huber, R. (1997). Reaction mechanism of *Escherichia coli* dihydrodipicolinate synthase investigated by x-ray crystallography and NMR spectroscopy. Biochemistry, 36, 24-33.

Borthwick, E. B., Connell, S. J., Tudor, D. W., Robins, D. J., Shneier, A., Abell, C. and Coggins, J. R. (1995). *Escherichia coli* dihydrodipicolinate synthase: characterization of the imine intermediate and the product of bromopyruvate treatment by electrospray mass spectrometry. Biochemical Journal, *305*, 521–524.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Gross-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., *et al*. (1998). Crystallography and NMR System (CNS): A new software system for macromolecular structure determination. Acta Crystallographica D54, 905-921.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallographica, *D50*, 760-763.

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java Alignment Editor. Bioinformatics, *20*, 426-427.

Coulter, C. V., Gerrard, J. A., Kraunsoe, J. A. E., Moore, D. J., and Pratt, A. J. (1999). *Escherichia coli* Dihydrodipicolinate Synthase and Dihydropicolinate Reductase: Kinetic and Inhibition Studies of Two Putative Herbicide Targets. Pesticide Science, *55*, 887-895.

Cox, R. J., Sutherland, A., and Vederas, J. C. (2000). Bacterial Diaminopimelate Metabolism as a Target for Antibiotic Design. Bioorganic and Medicinal Chemistry, *8*, 843-871.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., and McPherson, A. (1994). Screening and optimization stratagies for macromolecular crystal growth, Acta Crystallographica *D50*, 414-423.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.sourceforge.net. Retrieved February 21, 2008.

Dixon, R. O. (1969). Rhizobia (with particular reference to relationships with host plants). Annual Review of Microbiology, *23*, 137–158.

Dobson, R. C. J., Valegard, K., and Gerrard, J. A. (2004a). The Crystal Structure of Three Site-directed Mutants of *Escherichia coli* Dihydrodipicolinate Synthase: Further Evidence for a Catalytic Triad. Journal of Molecular Biology, *338*, 329-339.

Dobson, R. C. J., Gerrard, J. A., and Pearce, F. G. (2004b). Dihydrodipicolinate synthase is not inhibited by its substrate (*S*)-aspartate-β-semialdehyde. Biochemical Journal, *377*, 757-762.

Dobson, R. C. J., Devenish, S. R., Turner, L. A., Clifford, V. R., Pearce, F. G., Jameson, G. B., and Gerrard, J. A. (2005a). Role of Arginine 138 in the Catalysis and Regulation of *Escherichia coli* Dihydrodipicolinate Synthase. Biochemistry, *44*, 13007-13013.

Dobson, R. C. J., Griffin, M. D. W., Jameson, G. B. and Gerrard, J. A. (2005b). The crystal structures of native and (*S*)-lysine-bound dihydrodipicolinate synthase from *Escherichia coli* with improved resolution show new features of biological significance. Acta Crystallographica, *D61*, 1116-1124.

Emsley, P. and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. Acta Cystallographica, *D60*, 2126-2132.

Holm, L. and Park, J. (2000). DaliLite workbench for protein structure comparison. Bioinformatics, *16*, 566-567.

Hutton, C. A., Southwood, T. J., and Turner, J. J. (2003). Inhibitors of Lysine Biosynthesis as Antibacterial Agents. Mini-reviews in Medicinal Chemistry, *3*, 115-127

Jancarik, J. and Kim, S. H. (1991). Sparce matrix sampling: a screening method for crystallization of proteins. Journal of Applied Crystallography, *24*, 409-411.

Jeffery, P. (2006). http://xray0.princeton.edu/~phil/Faculty/Guides/xrayDataCollection.html Retrieved January 3, 2008.

Joint Center for Structural Genomics (JCSG). Crystal Structure of Dihydrodipicolinate Synthase (TM1521) from *Thermotoga maritima* at 1.80 Å Resolution. (1O5K).

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography, *26*, 795-800.

Kefala, G., Panjikar, S., Janowski, R., and Weiss, M. S. (2008) The Crystal Structure of Dihydrodipicolinate Synthase from *Mycobacterium tuberculosis*. Biochem. J., 411, 351-360.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature *181*, 662-666.

Kumarevel, T. S., Karthe, P., Kuramitsu, S., Yokoyama, S. (2007). Crystal Structure of Dihydrodipicolinate Synthase from Aquifex aeolicus. (2EHH).

Laber, B., Gomis-Rüth, R. X., Romão, M. J., and Huber, R. (1992). *Escherichia coli* Dihydrodipicolinate Synthase. Identification of the Active Site and Crystallization. Biochemical Journal, *288*, 691-695.

Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C. and Marcuccio, S. M. (1997). Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminate lyase. Journal of Molecular Biology *266*, 381-399.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, *26*.

Lin, A., Mu, J., Yang, J. and Roach, P. J. (1999). Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. Archives of Biochemistry and Biophysics, *363*, 163–170.

Mallery, C. (2006). http://fig.cox.miami.edu/~cmallery/150/gene/sf13x1box.jpg Retrieved November 16, 2007.

McCoy, A. J. (2005). http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phases.html. Retrieved November 16, 2007

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007). *Phaser* crystallographic software. Journal of Applied Crystallgraphy, *40*, 658-674.

Mirwaldt, C., Korndorfer, I., and Huber, R. (1995). The Crystal Structure of Dihydrodipicolinate Synthase from *Escherichia coli* at 2.5 Å Resolution. Journal of Molecular Biology, *246*, 227-239.

Murshudov, G. N.,Vagin, A .A. and Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallographica, *D53*, 240-255.

Phenix, C. P. (2007). Investigation of MosA, a protein implicated in rhizopine Biosynthesis. Ph.D. Thesis. University of Saskatchewan. Saskatoon, SK. Canada.

Rao, J. P., Grzemski, W. and Murphy, P.J. (1995). *Rhizobium meliloti* lacking *mosA* synthesizes the rhizopine scyllo- inosamine in place of 3-O-methyl-scyllo-inosamine. Microbiology, *141*, 1683 – 1690.

Ramakrishnan, C., and Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. Biophysical Journal, *5*, 909-933.

Reardon, D. and Farber, G. K. (1995). The structure and evolution of alpha/beta barrel proteins. FASEB J., 9, 497–503.

Roberts, S. J., Morris, J. C., Dobson, R. C. J. and Gerrard, J. A. (2003). The preparation of (*S*)-aspartate semi-aldehyde appropriate for use in biochemical studies. Bioorganic and Medicinal Chemistry Letters, *13*, 265–267.

Roussel, A. and Cambillau, C. (1990). TURBO-FRODO: A New Program for Protein Crystallography and Modelling. Acta Crystallographica, *A46*, C66.

Scapin, G and Blanchard, J. S. (1998). Enzymology of bacterial lysine biosynthesis. Advances in Enzymology and Related Areas of Molecular Biology, *72*, 279-324.

Tam, P. H., Phenix, C. P. and Palmer, D. R. J. (2004). MosA, a protein implicated in rhizopine biosynthesis in Sinorhizobium meliloti L5-30, is a dihydrodipicolinate synthase. Journal of Molecular Biology, *335*, 393-397.

Yugari, Y. and Gilvarg, C. (1965). The condensation step in diaminopimelate synthesis. Journal of Biological Chemistry, *240*, 4710-4716.