Abstract

Title of Document:	STRUCTURE FUNCTION DIVERSITY WITHIN THE PHOSPHOENOLPYRUVATE MUTASE / ISOCITRATE LYASE SUPERFAMILY AS REVEALED BY THE ENZYMES OXALOACETATE DECARBOXYLASE AND 2,3-DIMETHYLMALATE LYASE
	Buvaneswari Coimbatore Narayanan, Doctor of Philosophy, 2008
Directed By:	Professor, Osnat Herzberg, Molecular and cellular biology program

Two members of the phosphoenolpyruvate mutase (PEPM) / isocitrate lyase (ICL) superfamily were investigated to study their structure-function relationships and to identify sequence signatures that define a particular function. The first enzyme (PA4872) was a protein of unknown function from *Pseudomonas aeruginosa*. The second enzyme from *Aspergillus niger* (An07g08390) was thought to be an oxaloacetate acetyl hydrolase (OAH) because of its high sequence identity (~60%) to an enzyme with confirmed OAH activity.

The X-ray crystal structure determination of PA4872 revealed unique features that guided the design of biochemical experiments, which ultimately led to the discovery that the enzyme is an oxaloacetate decarboxylase (OAD). Two structures of An07g08390, one with bound Mg^{2+} and the second with bound Mn^{2+} and the inhibitor 3,3-difluorooxaloacetate, were determined. The functional studies demonstrated that although

the enzyme has OAH activity, it has a far better activity as a 2R,3S-dimethylmalate lyase (DMML). The active site structure of DMML indicated a proline residue (Pro240) as a marker of DMML function along with confirming the conserved locations of previously established signature residues for lyase activity.

OAD is the founding member of a family within the PEPM / ICL superfamily and thus defines the function of the remaining family members. However, the biological context in which OAD functions remains unknown. DMML is known to function in the nicotinate catabolism pathway but not all the members of the pathway are present in *A. niger*. Transcriptome analysis suggests that the DMML encoding gene is under carbon catabolite repression but the pathway in which the enzyme functions has not yet been identified.

STRUCTURE FUNCTION DIVERSITY WITHIN THE PHOSPHOENOLPYRUVATE MUTASE / ISOCITRATE LYASE SUPERFAMILY AS REVEALED BY THE ENZYMES OXALOACETATE DECARBOXYLASE AND 2,3-DIMETHYLMALATE LYASE

By

Buvaneswari Coimbatore Narayanan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Osnat Herzberg, Chair
Professor Marco Colombini
Associate Professor James Culver
Professor John Orban
Professor Steven Rokita

Preface

Declaration of author's intent to use own previously published text.

The main text, figures and figure legends for Chapter 3 and Chapter 4 were used, and only modified for readability/formatting and to avoid repetition within the context of this dissertation. Full citations for the two chapters are given below.

Citation and acknowledgements for chapter 3:

Buvaneswari C. Narayanan, Weiling Niu, Ying Han, Jiwen Zou, Patrick Mariano, Debra Dunaway-Mariano and Osnat Herzberg (2008). "Structure and function of PA4872 from Pseudomonas aeruginosa, a novel class of oxaloacetate decarboxylase from the PEP mutase/isocitrate lyase superfamily" Biochemistry **47**(1): 167-82.

B. C. N performed protein structure determination and W. N carried out kinetic determinations.

Citation and acknowledgements for chapter 4:

Buvaneswari C. Narayanan, Weiling Niu, Henk-Jan Joosten, Zhimin Li, Remko K. P. Kuipers, Peter .J. Schaap, Debra Dunaway-Mariano and Osnat Herzberg (2008). "Structure and function of 2,3Dimethylmalate lyase, a PEP mutase / isocitrate lyase superfamily member" In press, Journal of Molecular Biology.

B. C. N. determined the X-ray structures; W.N. carried-out the biochemical studies with help from Z. L.; H.-J. J. and R. K. P. K. cloned the gene from *A. niger*; P.J.S. carried out the DNA array analysis and *A. niger* growth experiments.

Dedication

To all my Gurus

Acknowledgements

I would like to express my most sincere gratitude to Prof. Osnat Herzberg, my research advisor, who has watched over and guided my development as a scientist. Her phenomenal insight in crystallography and enthusiasm in structure analysis have inspired me a countless number of times. The lessons I have learned from Prof. Herzberg, academic and otherwise, will be carried with me throughout my career.

I thank our collaborator Dr. Debra Dunaway-Mariano from the University of New Mexico. Her contribution to this work is immense and her inputs helped a great deal in structure function studies. I thank Dr. Peter .J. Schaapp from the Wageningen University, The Netherlands for the genetic and genomic context analysis of DMML. I would like to thank all of our coauthors: Weiling Niu, Ying Han, Jiwen Zou, Henk-Jan Joosten, Zhimin Li, and Remko. K. P. Kuipers.

Thanks to the members of my advisory committee: Drs. Marco Colombini, James Culver, John Orban and Steven Rokita for their guidance, teaching and support, these not being restricted to my committee meetings.

To all the members of Prof. Herzberg's lab, former and present especially Sadhana Pullalarevu, Kinlin Chao, Kap Lim, Celia Chen, Andrey Galkin, and Manish Chandra Pathak, you have all in some measure touched this research and I thank you. I express my special gratitude to Mark Willis who helped me immensely in my initial years learning crystallography. I thank Darwin Diaz for his efforts in CARB in maintaining and providing us with a great X-ray diffraction facility. I appreciate the help, support and friendship of Jane Ladner. To all of my friends at CARB and UMCP, you have made my graduate education more rich and enjoyable.

I thank my husband Prasanth, who experienced the ups and downs of graduate life with me and for giving me the love and space I needed to proceed in pursuit of my goals. I am also grateful for all his efforts in teaching probability theory and many more to me. My family has really been the foundation for all my achievements. I thank my parents Baby, Narayanan, my brother Vidyasankar and my sister in-law Hema for always being there for me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADP	Adenosine diphosphate
CAPSO	3-(cyclohexylamino)-2-hydroxy-1-propanesulfonic acid
CPEPM	carboxy phosphenolpyruvate mutase
DMM	Dimethyl malate
DMML	2,3-dimethylmalate lyase
DDT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
ICL	Isocitrate lyase
IPTG	Isopropyl-D-thiogalactopyranoside
LDH	Lactate dehydrogenase
MAD	Multiwavelength anomalous diffraction
MES	4-morpholineethanesulfonic acid
MICL	2-methylisocitrate lyase
MIR	Multiple isomorphous replacement
MOBH	5, 10-methylenetetrahydrofolate: 3-methyl-2-oxobutanoate hydroxymethyl transferase
MR	Molecular replacement
NADH	Nicotinamide adenine dinucleotide
OA	Oxaloacetate
OAD	Oxaloacetate decarboxylase
OAH	Oxaloacetate acetyl hydrolase
Panb	Ketopantoate hydroxymethyl transferase
PDB	Protein data bank
PDP	Petal death protein
PEP	Phosphoenolpyruvate
PEPM	Phosphenolpyruvate mutase
PPHA	Phosphonopyruvate hydrolase
PK	pyruvate kinase
TAPS	2-Hydroxy-1, 1-bis(hydroxymethyl)ethyl)amino]-1-propanesulfonic acid

Chapter 1 Introduction

Molecular recognition of substrates by enzymes depends on the active site structure. The size and shape of the molecule that is accommodated in the active site is dictated by the residue distribution lining the active site. In order to understand the function and mechanism of an enzyme, a deeper knowledge of the atomic details of the structure is important.

My research focus has been the structure/function study of the phosphoenolpyruvate mutase (PEPM) / isocitrate lyase (ICL) superfamily of proteins. This is an unusual superfamily because its members share high sequence identity and the same overall fold, yet they function in diverse biological contexts and carry out a wide range of biochemical functions. Structural studies of several PEPM/ICL superfamily members using X-Ray crystallography have already been carried out in our laboratory to gain insight into the structural determinants that define diversification of function and to identify signature residues that help in the assignment of function of new sequences that belong to the PEPM/ICL superfamily. My research concentrated on two of the superfamily members, oxaloacetate decarboxylase (OAD) and 2, 3–dimethylmalate lyase (DMML).

In the studies of OAD and DMML the structural work including crystallization, X-ray diffraction data collection, structure determination, refinement, validation and structure analysis were carried out by myself in the laboratory of Dr. Osnat Herzberg at the Center for Advanced Research in Biotechnology and the biochemical and enzymology work were carried out in the laboratory of Dr. Debra Dunaway-Mariano at the University of New Mexico. The molecular architecture revealed by the three

dimensional structures helped develop hypothesis about the catalytic mechanism and substrate specificity which were then tested by site directed mutagenesis. An introduction to the PEPM/ICL superfamily is covered in this chapter. An overview of the methodology (X-Ray-Crystallography) is described briefly in the following chapter.

1.1 The PEPM/ICL superfamily

The PEPM/ICL superfamily consists of enzymes that catalyze a C-C bond formation or cleavage and those that catalyze a P-C bond formation or cleavage reactions in α-oxycarboxylate substrates (Fig. 1.1). The known C-C bond catalyzing enzymes are isocitrate lyase (ICL) (Vanni, Giachetti et al. 1990), 2- methylisocitrate lyase (MICL) (Horswill and Escalante-Semerena 1999), petal death protein (PDP) (Lu, Feng et al. 2005; Teplyakov, Liu et al. 2005), DMML, OAD, oxaloacetate acetyl hydrolase (OAH) (Han, Joosten et al. 2007) and ketopantoate hydroxymethyl transferase (PanB) (Jones, Brook et al. 1993). Enzymes acting on a P-C bond are PEPM (Seidel, Freeman et al. 1988), carboxy phosphoenolpyruvate mutase (CPEPM) (Pollack, Freeman et al. 1992), and phosphonopyruvate hydrolase (PPHA) (Kulakova, Wisdom et al. 2003). The chemical reactions proceed via an oxyanion intermediate and/or transition state with the aid of a divalent metal cofactor (Mg^{2+} or Mn^{2+}). The lyase branch of the superfamily (ICL, MICL, OAH, PDP) catalyzes the cleavage of the C(2)-C(3) bond of various α-hydroxyacid substrates to form a α-oxyanioncarboxylate intermediate/transition state and an aci-carboxylate anion intermediate/transition state (Liu, Lu et al. 2005). In contrast, the known enzymes of the P-C bond forming/cleaving branch (PEPM, CPEPM and PPHA) catalyze phosphoryl transfer in an α-ketocarboxylate substrate via an enolate

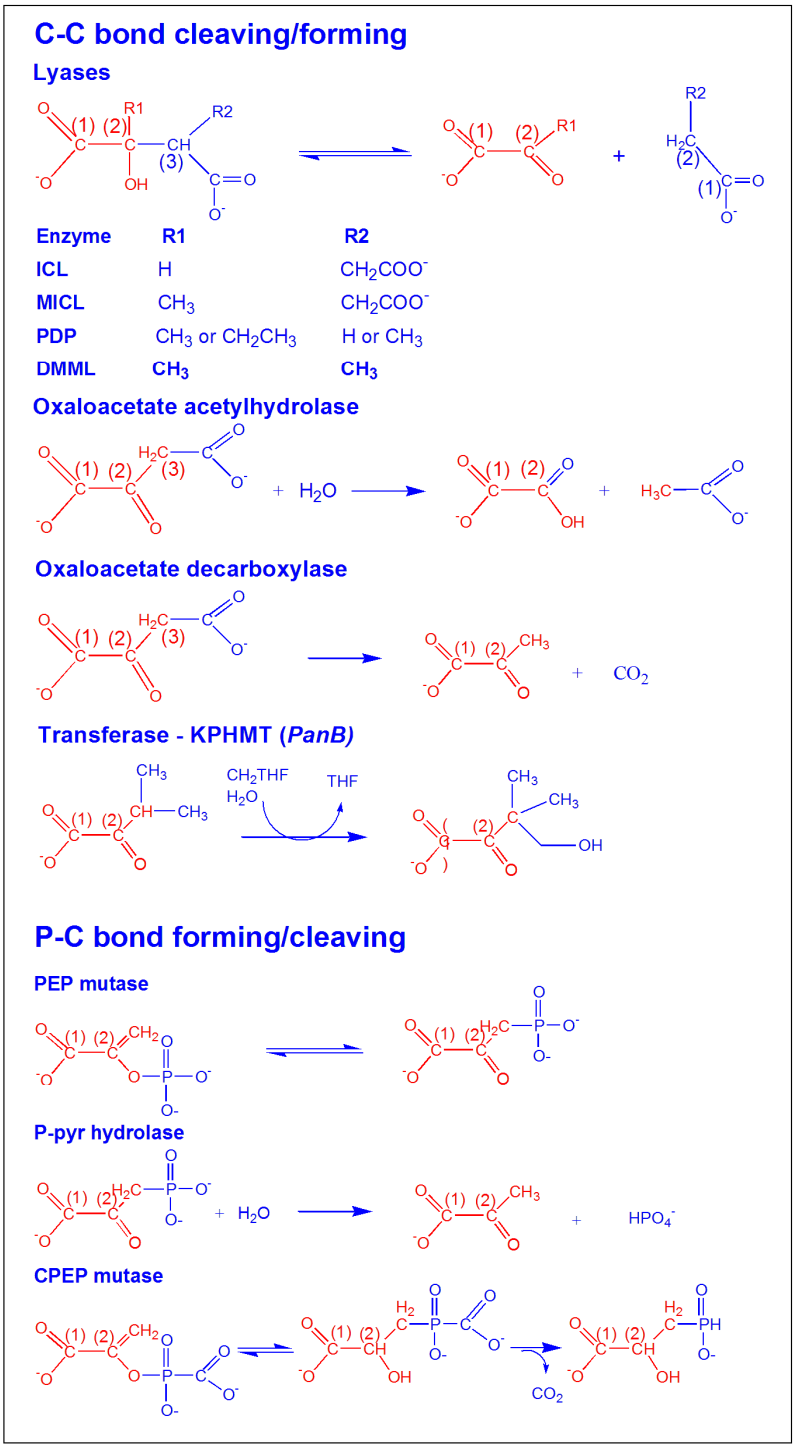


Figure 1.1: *Reactions catalysed by the PEPM/ICL superfamily. Conserved pyruvyl/glyoxyl moieties in the substrates and products are highlighted in red.*

intermediate/transition state (Huang, Li et al. 1999; Chen, Han et al. 2006). The stabilization of these intermediates form the core chemistry of these enzymes (Lu, Feng et al. 2005). Our work on OAD revealed that this enzyme also involves an enolate intermediate (Narayanan, Niu et al. 2008).

In addition to the wide range of biochemical functions, the PEPM / ICL superfamily enzymes function in diverse range of biological processes spanning primary and specialized secondary metabolisms. For example, the C-C bond cleaving activity of the highly specific enzyme, ICL, is utilized in the glyoxylate pathway (Vanni, Giachetti et al. 1990) and that of MICL is utilized in the methylcitrate cycle (McFadden, Rose et al. 1972; Brock, Darley et al. 2001). In contrast, PDP exhibits broad substrate specificity and is associated with programmed petal death, where it may be used to accumulate nutrients for seed development (Lu, Feng et al. 2005; Teplyakov, Liu et al. 2005). The enzymes that act on P-C bonds, PEPM and carboxyPEPM, function in phosphonate biosynthesis. Whereas, PPHA functions in phosphonate degradation.

Members of the PEPM/ICL superfamily usually oligomerize into tetramers consisting of dimer of dimers (Fig 1.2) except in PanB where the dimers form decamers (pentamers of dimers) (Chaudhuri, Sawaya et al. 2003; Schmitzberger, Smith et al. 2003). A recently published paper on a new member of unknown function from the superfamily, DFA0005, reported that in this case the dimer is the physiologically relevant assembly (Liao, Chin et al. 2008). The basic dimeric building unit associate similarly regardless of the overall oligomeric state. The overall fold characteristic to the PEPM/ICL superfamily is an α/β-barrel fold where the eighth α-helix of one barrel extends out to pack against the barrel of the partner protomer (except for one PanB structure from E.coli

(Schmitzberger, Smith et al. 2003)). After swapping, the eighth α-helix ensuing chain forms a helix- loop- helix C-terminus that covers the active site of the partner molecule. The structure of OAD that I determined provides a new variation in which the C-terminus switch the direction back to its own active site. The active site is located at the C-terminal edge of the barrel's β sheets. The cavity is gated by a capping loop that undergoes conformational changes, yielding an open or closed active site. The open conformation exposes the active site to bulk solvent and enables substrate binding and product release. Upon substrate binding, the loop closes to sequester the active site from the solvent. The gating loop closure brings key catalytic residues in close proximity to the substrate and protects the intermediate/transition states from solvent.

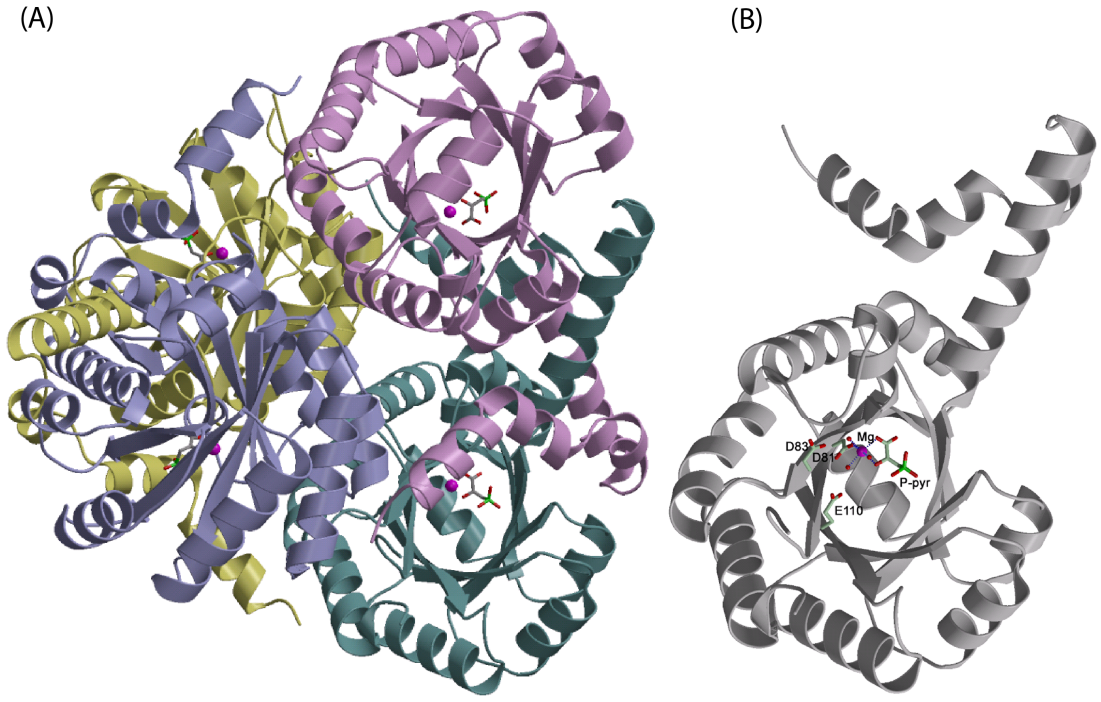


Figure 1.2: *(A) Tetramer structure of PPHA. The dimer association showing helix swapping is seen between the monomers colored green and purple (B) Monomer of PPHA (adapted from (Chen, Han et al. 2006)) showing* Mg^{2+} *as magenta sphere and phosphonopyruvate (stick representation) in the active site.*

The catalytic scaffold provides a binding pocket for the α-oxycarboxylate substrate that includes residues engaged in hydrogen bond formation with the C(1) carboxylate, and residues that bind the Mg^{2+} cofactor directly or *via* water molecules (Fig. 1.3). The presence of Mg^{2+} or Mn^{2+} is crucial for effective catalysis in all PEPM / ICL superfamily enzymes and its coordination geometry and mode of engaging both protein residues and substrate are conserved.

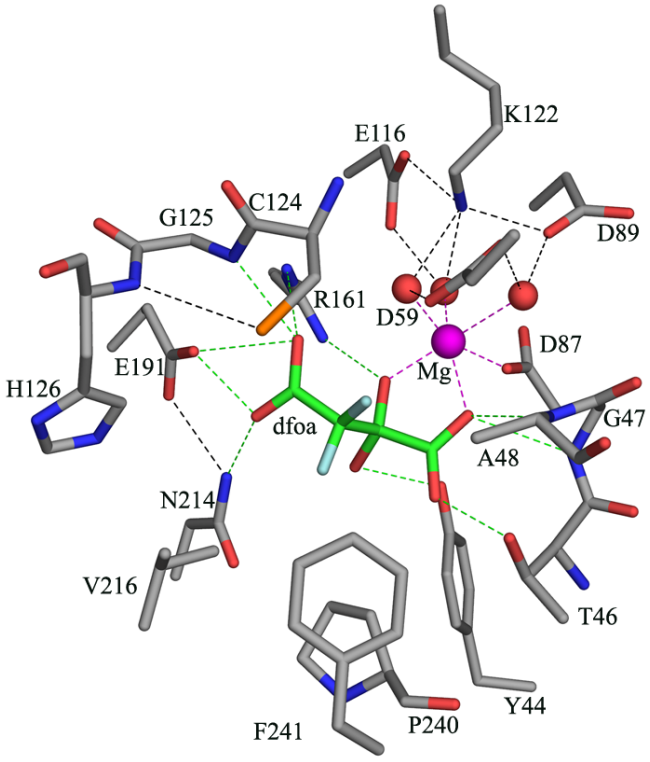


***Figure 1.3**: Active site of DMML*

The sequence alignment of the PEPM / ICL superfamily members shows that a number of residues are stringently conserved among these proteins. The cofactors coordinate directly the Asp87 side chain (all residue numbering are according to DMML structure throughout this section unless indicated otherwise) and indirectly via three water molecules, the side chain carboxylic groups of Asp59, Asp89 and Glu116 (depicted for DMML in figure 1.3). These Mg^{2+}/Mn^{2+} coordinating residues are invariant as

highlighted in red in figure 1.4. The Mg^{2+} cofactor coordinates one oxygen atom of the substrate C(1) carboxylate group and the C(2) α-oxygen atom, thereby serving as an electron sink in the stabilization of the oxyanion intermediate and/or transition state. The conserved capping loop residue Lys122 (except for OAD described in chapter 3) involved in loop opening and closure (Liu, Lu et al. 2004) is stationed in the vicinity of Mg^{2+}/Mn^{2+} in the closed gating loop conformation with two water molecules in between to reduce electrostatic repulsion.

Some of the residues that interact with the substrate are also conserved. Because the various substrates share a common pyruvyl/glyoxyl moiety, oxalate, an anolog of the shared pyruvyl adduct, is an inhibitor of all PEPM/ICL superfamily members. The residues that interact with the conserved pyruvyl/glyoxyl moiety are highlighted in blue in figure 1.4. The inhibitor C(1)OO- carboxylate interacts with the Thr46 hydroxyl group (serine in the other lyases, PEP and PPHA), and with an oxyanion hole formed by the backbone amides of Gly47 and Ala48 which are located at the N-terminus of a conserved short α-helix. One of the inhibitor's C(2) oxygen atoms engages in an ionic interaction with the guanidinium group of Arg161 and is stationed in an oxyanion hole formed by Mn^{2+} and Arg161. Arg161 is an invariant residue that plays a key catalytic role.

In addition to the residues that are conserved throughout the PEPM/ICL superfamily, there are signature residues that discriminates functional families within the superfamily and underscores the differences that makes each member uniquely tailored for its particular function. For example, the gating loops of the lyases, OAH, and CPEPM, carry a catalytic cysteine, whereas PEPM and PPHA contain asparagine and threonine, respectively.

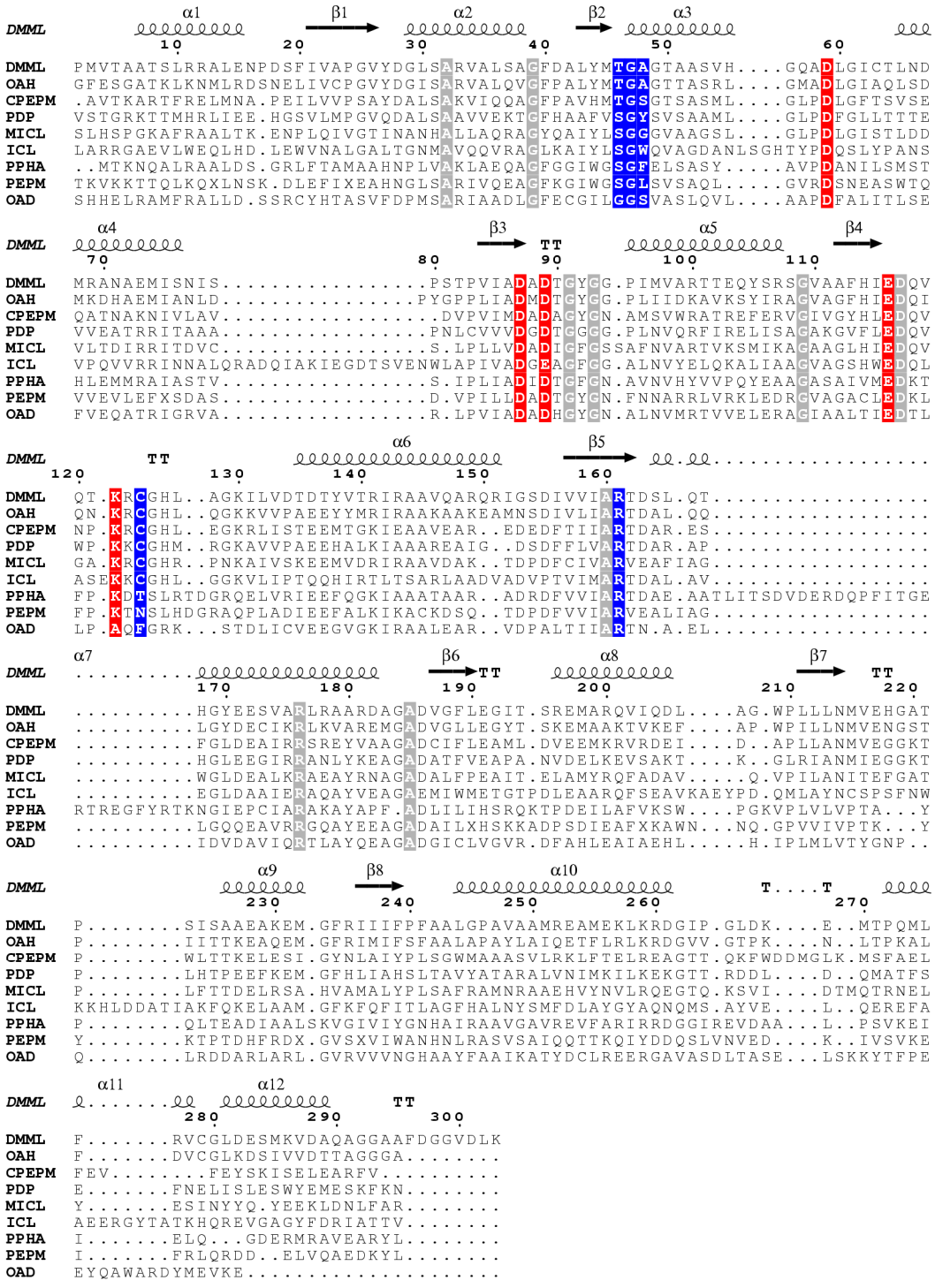


Figure 1.4: *Sequence alignment of the PEPM/ICL superfamily members. Conserved Mg^{2+}/Mn^{2+} coordinating residues are highlighted in red, the substrate/inhibitor coordinating residues are colored blue and all other invariant residues are colored gray.*

The sequence identity between members of the superfamily with different functions may be as low as ~20% and as high as ~60%. Low sequence identity is often encountered within a superfamily; however, it is quite unusual for enzymes with sequence identity of 40% (viz. PDP and carboxyPEPM) or 57% (viz. OAH and DMML) to have different functions. This underscores the blurring of boundaries between function families and the uncertainties of database annotations. A phylogeny tree generated in our laboratory using enzymes of known structures and/or confirmed function as seed for multiple sequence alignment shows clear separation between several function groups (Fig. 1.5), yet it also reveals a region of ambiguity of function and regions where the function has not yet been identified (Liu, Lu et al. 2005).

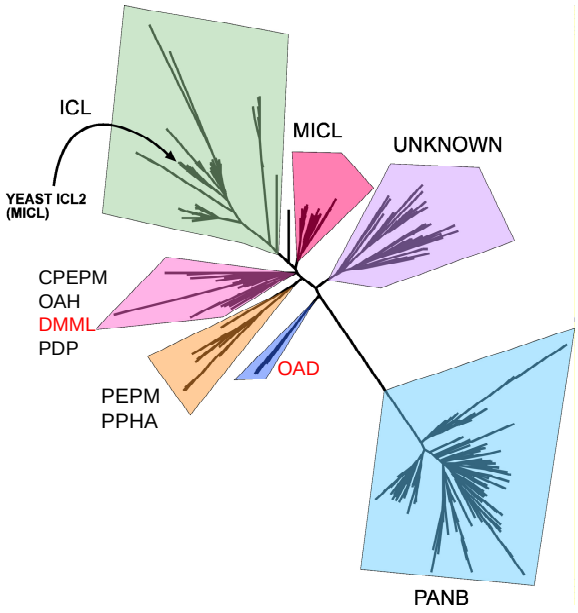


***Figure 1.5**: Phylogeny tree of the PEPM/ICL superfamily. Adapted from (Liu, Lu et al. 2005)*

The ICL and MICL enzymes form two well separated groups but the ICL2 type MICLs are contained within the ICL family cluster. There are differences in substrate binding motifs between ICL and MICL, which is evident from the structures of the two enzymes (Liu, Lu et al. 2005). The P-C bond breaking family including PEPM and PPHA

is also well separated from other families. The cluster comprising the more distant transferases (PanB) which do not possess the gating loop, oligomerizes uniquely into decamers, and differ in several key active site residues, is well separated from the rest of the enzyme clusters.

The phylogeny tree revealed two distinct branches comprising proteins of unknown functions. One branch includes the *Pseudomonas aeruginosa* PA4872, the first protein I studied, which subsequently was shown by us to be a novel oxaloacetate decarboxylase (Narayanan, Niu et al. 2008). The crystal structure determination was crucial to function assignment because it provided the first clue that the catalytic mechanism of the enzyme involves an enolate intermediate and ruled out lyase activity involving a carbanion intermediate (Liu, Lu et al. 2005). Chapter 3 provides a detailed account of the studies on this enzyme. The function of the second family cluster in the phylogeny tree remains unknown despite the recent structure report of one member (Liao, Chin et al. 2008), which contains an intrinsically bound ligand interpreted by the authors as α-ketoglutarate.

The family adjacent to MICL in the phylogeny tree includes enzymes with different functions despite the high sequence identity they share: CPEPM, PDP, and OAH. The second protein that I studied, 2,3-dimethylmalate lyase belongs to this clade. Before we began our studies, the protein was thought to be OAH because it exhibits ~60% sequence identity to a confirmed OAH. Chapter 4 provides a complete account of these studies. Because of the diverse activities within this phylogeny clade, it may include members that have different, yet to be discovered functions, which is underscored by the discovery of DMML.

Chapter 2 X-Ray crystallography: Methodology

X-ray crystallographic methods were employed to determine the protein structures described in this thesis. The experimental and computational techniques used were summarized briefly in chapter 3 and 4 to enable other researchers reproduce the results. This chapter describes the procedures I employed in my studies in more detail.

2.1 Crystallization of macromolecules

Despite much research to improve the predictability of protein crystal growth, the method remains empirical. However, extensive experience that has been accumulated over the years has led to the development of techniques that accelerate and automate the crystallization process. In my studies, I used the vapor diffusion method for crystallization. In the vapor diffusion method, small volumes of precipitant solution and protein are mixed together and the drop is equilibrated against a larger reservoir of solution containing just the precipitant solution. The method can be employed either with a hanging drop or a sitting drop setup. In the hanging drop method a drop containing the mixture of protein sample and precipitant solution is placed on a siliconized coverslip, flipped over to cover and seal the reservoir containing the precipitant solution. In the sitting drop method the crystallization solution is placed in the reservoir's well, containing a bridge that holds the drop of the protein/precipitant mixture (Fig 2.1). The sitting drop arrangement is used in crystallization robots usually in a 96-well format. In most experiments the reservoir contains higher concentration of precipitant solution and equilibration leads to the increase in concentration of the protein solution, which leads to super saturation and sometimes nucleation and crystal growth.

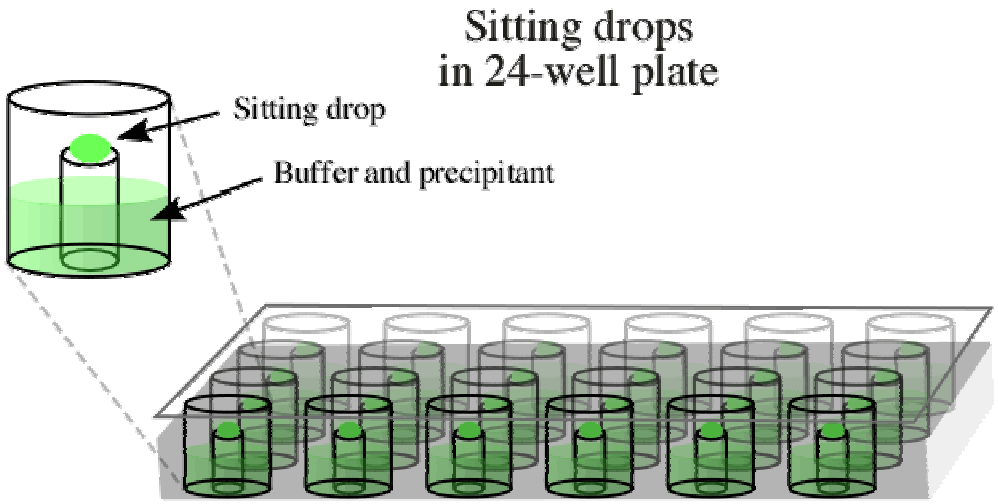


***Figure 2.1**: A typical sitting drop crystallization setup.*
Source:http://www.usm.maine.edu/chy/manuals/BiochemLab/Text/HdtXtals/images/wellplate.gif

In practice, initial screens of various preformulated conditions were tested employing both the hanging drops and sitting drops vapor diffusion methods using commercially available screens. Once a condition was identified that displayed crystals or the potential to develop crystals, crystal growth was optimized by fine sampling of various parameters that affect crystallization such as the precipitant concentration (ammonium sulphate, various polyethylene glycols), pH of the solution, the temperature at which the crystallization trays were modified, and also the presence/absence of salt and metal additives. Figure 2 shows the crystals of OAD (2A) and DMML (2B) useful for X-ray diffraction.

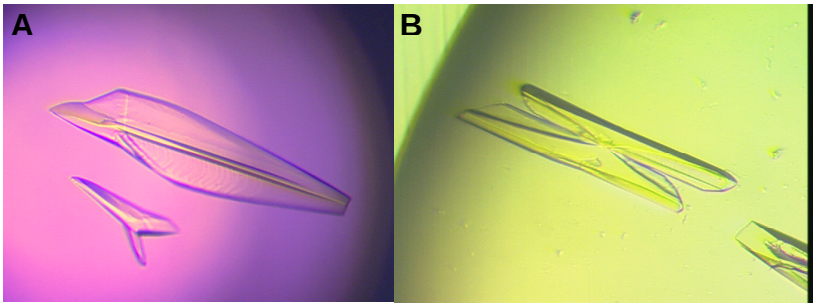


***Figure 2.2:** (A) OAD – Mg^{2+} - oxalate crystals (B) DMML- Mn^{2+} - 3,3-difluorooxaloacetate crystals*

2.2 Data collection

X-rays generated from a copper source (CuKα wavelength is 1.5418Å) were used on the home facility, but with the advent of tunable synchrotron X-rays, we exploited other wavelengths. To obtain an accurate structure, X-ray diffraction data collection should optimally cover the entire reciprocal space and extend to the maximum possible resolution characteristic of the particular crystal. The most efficient and universally employed method to acquire the diffraction data is the rotation or oscillation method. In this method the diffraction images (Fig. 2.3) are collected by rotating the crystal around a fixed axis in small continuous steps (0.25° - 0.5°) and the diffraction intensities are recorded on a detector.

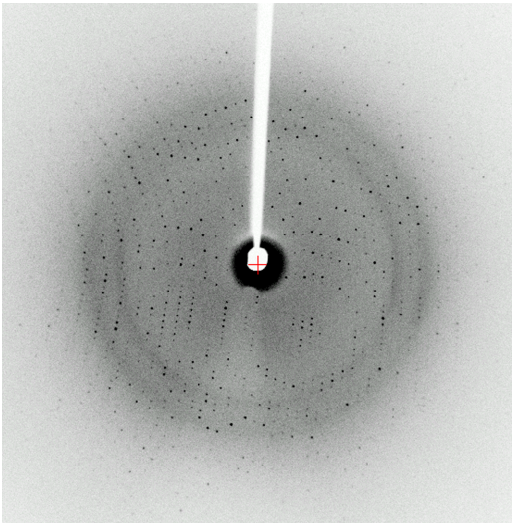


***Figure 2.3**: X-ray diffraction image of DMML- Mn^{2+}- inhibitor complex crystal collected at 100K on the CARB X-ray facility: Micromax 007 generator and R-AXIS IV^{++} image plate detector (Rigaku MSC)*

To reduce radiation damage, X-ray data collection is carried out at ~100K using a gaseous stream of cooled nitrogen. The crystals are first coated or soaked with cryoprotectants such as glycerol, ethylene glycol, or oil and then flash cooled. Optimal strategies for data collection are influenced by factors such as diffraction quality, cell

dimensions, crystal symmetry and crystal orientation relative to the incident X-ray beam. These parameters define the choice of exposure time and crystal to detector distance to obtain well separated diffraction spots in all crystal orientations and to maximize the signal to noise of the intensity measurements.

Diffraction data for the DMML and wild-type PA4872 crystals were collected on the home facility consisting of Osmic Max-Flux monochromated X-rays generated by a Rigaku Micro Max 007 rotating anode equipped with a Rigaku R-AXIS IV^{++} detector. X-ray diffraction data for the SeMet-containing Pa4872 protein crystal were collected in the synchrotron facility on the IMCA-CAT 17-ID beam line at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) utilizing three wavelengths for a MAD experiment (discussed later).

2.3 Data processing

The X-ray diffraction images are used to obtain a set of indices (*hkl*) and their intensities *I(hkl)* along with the intensity errors *σI(hkl).* Initially, the diffraction images are indexed by identifying a limited number of diffraction spots. The spot positions are used to deduce the crystal unit cell parameters and orientation with respect to the laboratory frame of reference by a least squares procedure. The lattice with highest symmetry that fits the reference spots with the lowest residual is selected. This is followed by the refinement of crystal and detector parameters and prediction of the positions of all diffraction spots in all images. The intensity of each of the diffracted spots is then integrated and corrected for background noise and systematic errors (polarization and Lorentz factor). The images are scaled using reflections that are measured multiple times to maximize the agreement between their intensity values. The structure factor

amplitudes are derived from the relationship: $F = \sqrt{I}$ where, F is the structure factor amplitude and I, the intensity. The structure factor expression is:

$$\overline{F}_{hkl} = V \sum_{j=1}^{N} f_j \exp 2\pi i(\overline{x}_j . \overline{h})$$

where, V is the volume of the unit cell, f_j is the scattering factor of atom j, $\overline{h} = h, k, l$ are the miller indices of a reflection, $\overline{x} = x, y, z$ are the fraction coordinates of atom j in real space and the summation is over all atoms 1 to N.

2.4 The Phase Problem

The measured intensities provide the structure factor amplitudes. However, structure factors are vectors and the phases that are essential for calculating the electron density map are lost when deriving the amplitudes from the intensities. Hence, the phases must be extracted through other means. Various methods are employed to obtain phases including direct methods, molecular replacement (MR), multiple isomorphous replacement (MIR) and multiwavelength anomalous diffraction (MAD). When there are problems of noise or unisomorphism, more than one of these methods may be combined to obtain the best possible phases. The following sections describe the MR and MAD methods as practiced in my studies.

2.4.1 Molecular replacement (MR)

MR is used to solve the phase problem when a closely related structure is available. Usually, but not always, this entails high sequence identity with the sequence of the protein of interest. The technique is now widely used because the PDB often contains a structure that can serve as a MR search model. In MR, the model structure is

oriented and positioned in the crystal cell so that the calculated diffraction best matches the observed diffraction. This is done in two separate steps: the rotation search and the translation search (Rossmann and Blow 1962) and in practice the calculations are carried out in reciprocal space. Once the position of the molecule in the cell is defined, the initial map is calculated using the experimentally observed amplitudes and the phases generated using the model structure as it is placed in the cell. The resulting electron density map is then used to improve the model by modifying incorrectly positioned amino acid residues.

The MR program I employed to obtain phases for the two DMML structures was PHASER (McCoy, Grosse-Kunstleve et al. 2007). The program incorporates a maximum-likelihood treatment that calculates the probability that the observations will be made given a particular model. The model with the highest likelihood is then chosen as the one that best accounts for the observations (Read 2001).

2.4.2 Multiwavelength anomalous diffraction

The Multiwavelength anomalous diffraction (MAD) technique is employed when a heavy atom with large anomalous signal is present in the crystal, which is often implemented in protein crystallography by incorporating selenium in the place of sulphur in methionine residues. In MAD, the wavelength is tuned to the absorption edge of the heavy atom to maximize the anomalous scattering signal arising from the electrons around the heavy atoms that absorb the incident X-rays and diffract with a phase difference (Hendrickson, Smith et al. 1985). Because of the phase difference, the amplitudes of the two reflections that form a Bijvoet pair are not equal to one another, which facilitate phase determination. The initial phases are obtained by first locating the positions of the heavy atoms and calculating the phases by a derivation analogous to that

used in MIR. It should be noted that there are actually two possibilities for the initial phases +φ and –φ. This phase ambiguity is solved by collecting an additional dataset at a wavelength close to the absorption edge of the heavy atom in the crystal where the anomalous signal is maximized. In solving the PA4872 (OAD) structure, I used the program SHELXD (Schneider and Sheldrick 2002) to locate the heavy atom sites and the program MLPHARE (Otwinowski 1991) to refine those sites and calculate the initial phases. The phases were further improved by density modification methods (Cowtan 1994), and these improved phases were used to calculate the initial electron density map.

2.5 Electron density calculation and model building

The electron density is the Fourier transform of all the structure factors at position (x,y,z):

$$\rho(x,y,z) = \frac{1}{V}\sum_{\bar{h}=-\infty}^{+\infty} \bar{F}_{hkl} \exp -2\pi i(\bar{x}.\bar{h})$$

where, V is the volume of the unit cell, $\bar{F}_{hkl}$ is the structure factor (amplitude and phase), $\bar{h}$ is the miller indices (h,k,l), $\bar{x}$ is the fraction coordinates (x,y,z) and the summation is over $\bar{h}$ from -∞ to +∞. The coefficients $2F_o$-F_c and calculated phases are used to calculate the electron density map where F_o and F_c are the observed and calculated structure factors. The electron density map of a section of the OAD structure is shown in figure 2.4 (A).

An initial model is built into the calculated electron density map using interactive graphics workstation. Figure 2.4(B) depicts the OAD polypeptide segment that was built into the electron density map shown in figure 2.4(A). The initial model was built automatically for the OAD structure using the programs Solve/ Resolve (Terwilliger and

Berendzen 1999; Terwilliger 2003), whereas the DMML initial model was obtained from the MR solution and required only placement of sidechains and modifications in regions that differed from the search model. The manual model rebuilding for the OAD structures and DMML structures were carried out using the graphics programs "O" (Jones, Zou et al. 1991) and COOT (Emsley and Cowtan 2004), respectively.

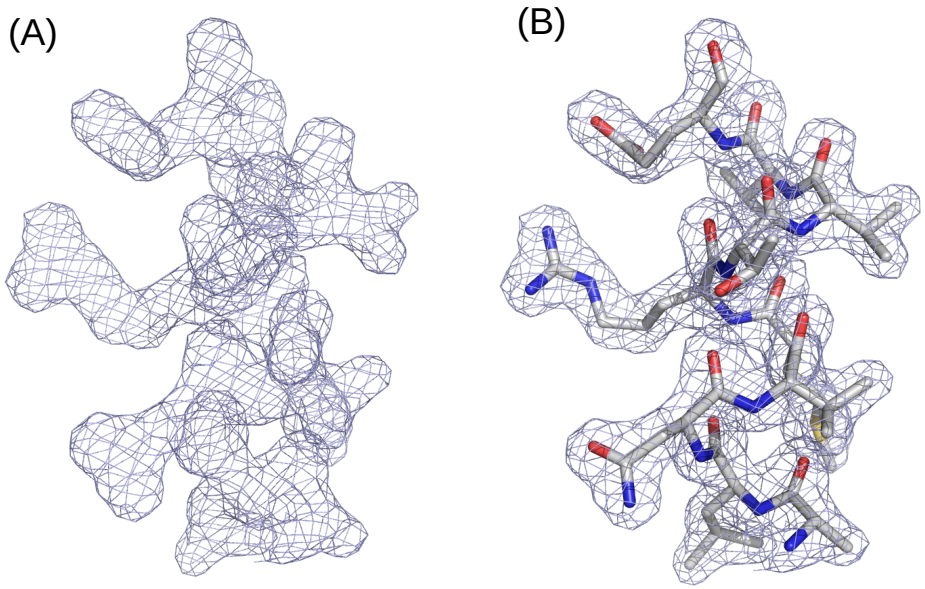


***Figure 2.4**: (A) The $2F_o$-F_c electron density map of OAD at 1.9Å resolution. The coefficients $2F_o$-F_c and calculated phases are used. (B) Polypeptide chain built into the electron density cage of OAD. The polypeptide chain in this region adopts an α-helical conformation*

2.6 Structure refinement

The initial model is usually inaccurate and needs to be improved using iterative steps of refinement and model modification on an interactive graphic workstation. The computer refinement program shifts atom positions and temperature factors to achieve better agreement between the observed and calculated structure factors. During the latter stages of refinement, water molecules, metal ions and other ligands (substrates/inhibitors, glycerol) are added into the electron density map. The resulting model is inspected

together with the new electron density map and a further round of refinement is carried out. This continues until the correlation between the observed and calculated structure factors is maximized. The agreement is measured by the *R*-factor and R_{free} values. *R*-factor is defined as:

$$R = \frac{\sum_{hkl} || \overline{F}_{hkl} |_{obs} - | \overline{F}_{hkl} |_{calc}|}{\sum_{hkl} | \overline{F}_{hkl} |_{obs}}$$

R_{free} is calculated from a subset (5-10%) of reflections that were not included in refinement. It is important to assure that the number of variables that are being refined (x,y,z, temperature factors, occupancies) does not exceed the number of observations, otherwise the problem is underdefined. Because of the limited diffraction resolution of most protein crystals, the number of reflections measured is almost always insufficient. Thus, geometrical restraints such as bond lengths, bond angles, chirality and planarity are included to increase the number of observations. The weight given to the geometrical restraints are calculated so that the root mean squares deviations are comparable to those found in small molecule structures determined at atomic resolution where no restraint are used during refinement. The refinement programs I used (CNS (Brunger, Adams et al. 1998) and REFMAC (Murshudov, Vagin et al. 1997)) both employed the maximum likelihood weighting where the weight is chosen according to the refinement target function and the current model.

2.7 Structure validation

Upon completion of the refinement, the quality of the structure is analysed. The program PROCHECK (Laskowski, Moss et al. 1993) was used to verify the stereo

chemical quality of the protein structures. The output consists of a detailed listing of parameters and their graphical representation, for each residue. These include the Ramachandran plot, deviation from bond angles and bond length, chirality checks, side chain dihedral angles and short contacts. The program compares and assesses the structure quality with respect to other structures at comparable resolution. Outstanding outlier regions are inspected to identify the source of the error and corrected. The refined structure (both the coordinates and the structure factor) is then submitted to the Protein Data Bank for release to the public after validation process.

Chapter 3 Structure and function of PA4872 from *Pseudomonas aeruginosa,* a novel oxaloacetate decarboxylase from the PEP mutase / isocitrate lyase superfamily

3.1 Abstract

Pseudomonas aeruginosa PA4872 was identified by sequence analysis as a structurally and functionally novel member of the PEP mutase/isocitrate lyase superfamily and therefore targeted for investigation. Substrate screens ruled out overlap with known catalytic functions of superfamily members. The crystal structure of PA4872 in complex with oxalate (a stable analog of the shared family α-oxyanion carboxylate intermediate/transition state) and Mg^{2+} was determined at 1.9 Å resolution. As with other PEP mutase/isocitrate lyase superfamily members, the protein assembles into a dimer of dimers with each subunit adopting an α/β barrel fold and two subunits swapping their barrel's C-terminal α-helices. Mg^{2+} and oxalate bind in the same manner as observed with other superfamily members. The active site gating loop, known to play a catalytic role in the PEP mutase and lyase branches of the superfamily, adopts an open conformation. The N^{ε} of His235, an invariant residue in the PA4872 sequence family, is oriented towards a C(2) oxygen of oxalate analogous to the C(3) of a pyruvyl moiety. Deuterium exchange into α-oxocarboxylate-containing compounds was confirmed by ^{1}H-NMR spectroscopy. Having ruled out known activities, the involvement of a pyruvate enolate intermediate suggested a decarboxylase activity of an α-oxocarboxylate substrate. Enzymatic assays led to the discovery that PA4872 decarboxylates oxaloacetate (k_{cat} = 7500 s^{-1} and K_m = 2.2 mM) and 3-methyloxaloacetate (k_{cat} = 250 s^{-1} and K_m = 0.63 mM). Genome context of

the fourteen sequence family members indicates that the enzyme is used by select group of Gram-negative bacteria to maintain cellular concentrations of bicarbonate and pyruvate; however the decarboxylation activity cannot be attributed to a pathway common to the various bacterial species.

3.1 Introduction

The studies described in this paper were carried out to determine the biochemical function of the hypothetical protein PA4872 from *Pseudomonas aeruginosa*. PA4872 is a member of the PEP mutase/isocitrate lyase (PEP mutase/isocitrate lyase) superfamily (Liu, Lu et al. 2005). Enzymes of the PEP mutase/isocitrate lyase superfamily catalyze either P-C or C-C bond formation/cleavage. The P-C bond modifying enzymes that have been characterized to date include PEP mutase, carboxyPEP mutase and phosphonopyruvate hydrolase (reactions 1-3 of Figure 3.1). Enzymes known to act on C-C bonds include 5, 10-methylenetetrahydrofolate:3-methyl-2-oxobutanoate hydroxymethyl transferase (MOBH), isocitrate lyase, 2-methylisocitrate lyase, the plant petal death protein (a broad specificity 2-alkyl malate lyase having an additional oxaloacetate hydrolase activity) and fungal oxaloacetate hydrolase (reactions 4-7 of Figure 3.1). A detailed introduction to the PEPM /ICL superfamily is given in Chapter 1.

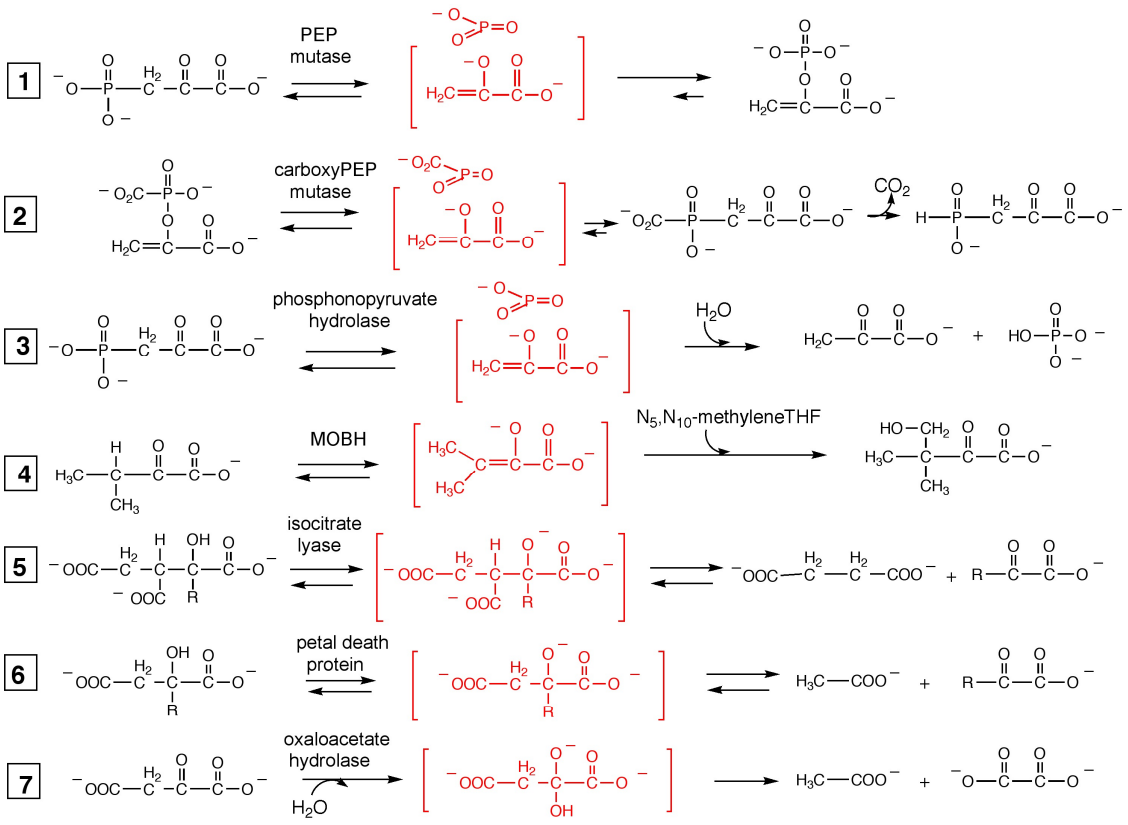


Figure 3.1*. The reactions catalyzed by the PEP mutase/isocitrate lyase superfamily. The reactions are depicted to occur stepwise with the formation of reaction intermediates (shown in brackets). The reactions might be concerted in which case the transition state will resemble the intermediate.*

To discover novel catalytic functions within the PEP mutase/isocitrate lyase superfamily we have focused our attention on the residue usage and the length of the gating loop possessed by members having unknown function. An alignment of family sequences identified PA4872 as a representative of a unique clad containing eight members at the time of analysis (now 14 members) and having a novel sequence for the active site gating loop (Liu, Lu et al. 2005). We set forth to discover the reaction catalyzed by PA4872 by employing a functional genomic approach that combines structure determination, activity screening and genome context neighbor analysis. Here we present the results, which show that PA4872 is an oxaloacetate decarboxylase, used

primarily by *Pseudomonads*. Inspection of genome contexts led us to speculate that the enzyme is employed by different bacterial species to maintain pyruvate and/or bicarbonate cellular levels in a variety of pathways.

3.3 Materials and methods

3.3.1 Preparation of recombinant wild-type PA4872

PA4872 was amplified from *P. aeruginosa* using *Pfu* Turbo DNA polymerase (Stratagene), genomic DNA, and oligonucleotide primers with restriction sites for NdeI and BamHI. The purified product was ligated to a *Nde*I- and *Bam*HI-cut pET3c vector (Novagen). The resulting clone, PA4872-pET3c, was used to transform *E. coli* BL21 (DE3) cells (Novagen). The gene sequence was verified by DNA sequencing carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine. Transformed *E. coli* cells were grown to OD_{600nm} ~ 0.8 at 25 °C in Luria broth (LB) media containing 50 μg/ml carbenicillin. Overexpression was induced with 0.4 mM isopropyl-D-thiogalactopyranoside (IPTG) at 25 ^{o}C. The cells were harvested by centrifugation at 6500 rpm (7,808*g*) for 15 min at 4 °C and suspended in 150 mL of lysis buffer (50 mM Na^{+}Hepes (pH 7.5), 5 mM DTT, 1 mM EDTA, 1 mM benzamide hydrochloride, 0.05 g/L trypsin inhibitor, 1 mM 1,10-phenanthroline, and 0.1 mM phenylmethylsulphonylfluoride). Cells were lysed using a French press and then centrifuged for 60 min 15 min at 4 °C. The protein of the supernatant was fractionated by $(NH_4)_2SO_4$ induced precipitation and harvested by centrifugation. The PA4872-containing fraction (40-60 % saturated $(NH_4)_2SO_4$) was dissolved in buffer A (50 mM triethanolamine (pH 7.5), 5 mM $MgCl_2$ and 0.5 mM DTT), and then dialyzed against 4 x 1 L buffer A at 4 ^{o}C before loading onto a 3.5 x 60 cm DEAE-cellulose

column. The column was washed with 1 L buffer A and then eluted with a 1.4 L linear gradient of KCl (0 – 0.8 M) in buffer A. The chromatography was monitored by measuring the absorbance of the eluted fractions at 280 nm and by carrying out SDS-PAGE analysis. The desired fractions (eluted at ~ 0.4 M KCl) were combined, mixed with $(NH_4)_2SO_4$ to 20% saturation, and loaded onto a 3 x 30 cm phenyl Sepharose column equilibrated with 20% saturated $(NH_4)_2SO_4$ in buffer A. The column was washed at 4 ^oC with 500 mL of $(NH_4)_2SO_4$ (20% saturation) in buffer A and then eluted with a 700 mL linear gradient of $(NH_4)_2SO_4$ (20% - 0% saturation) in buffer A. The desired fractions (eluted at 0% of ammonium sulfate) were combined, mixed with $(NH_4)_2SO_4$ to 25% saturation, and loaded onto a 3 x 30 cm butyl Sepharose column equilibrated with 25% saturated $(NH_4)_2SO_4$ in buffer A. The column was washed at 4 ^oC with 500 mL of 25% saturated $(NH_4)_2SO_4$ in buffer A and then eluted with a 800 mL linear gradient of $(NH_4)_2SO_4$ (25% - 0% saturation) in buffer A. The desired fractions (eluted at ~12% saturated $(NH_4)_2SO_4$) were combined and concentrated to 15 mL using an Amicon concentrator (Amicon), dialyzed against 4 × 1 L buffer A, and then concentrated to 35 mg/mL using MACROSEP 10K OMEGA (centrifugation at 6500 rpm). The purified protein (yield 4.4 mg/gm wet cells) was shown to be homogeneous by SDS-PAGE analysis and stored at – 80 ^oC. The molecular mass measured by TOF MS-ES mass spectrometry (University of New Mexico Mass Spectrometry Lab) is 31321 Da, identical to the protein's theoretical molecular mass.

3.3.2 Preparation of selenomethionine containing PA4872

For the preparation of selenomethionine (SeMet)-containing protein, *E. coli* BL21 (DE3) cells transformed with the PA4872-pET3c vector were grown in minimal media

containing SeMet instead of methionine. The overexpression and purification protocols used are the same as those employed for the production of the wild type protein.

3.3.3 Preparation of PA4872 site directed mutants

Site-directed mutagenesis was carried out by using the Quikchange method and Turbo*pfu* polymerase. The plasmid PA4872-pET3c was used as the template for the PCR amplification. The mutant proteins were purified using the same protocol as described for the wild-type recombinant enzyme. The homogeneity (>90%) of the purified mutant proteins was confirmed by SDS-PAGE analysis. Mutant protein yield: ~3 mg/gm of wet cells.

3.3.4 PA4872 crystallization and data collection

Wild-type PA4872 was crystallized in hanging drops by the vapor diffusion method at room temperature. The drops contained equal volumes of reservoir solution and 10 mg/mL protein in 5 mM $MgCl_2$, 5 mM oxalate and 10 mM Na^+Hepes (pH 7.0). The reservoir solution contained 18% polyethylene glycol 3350, 20% glycerol and 0.1 M MES (pH 6.0). Crystals appeared within two days. SeMet-containing PA4872 was crystallized in sitting drops by the vapor diffusion method at room temperature. Aliquots of 10 mg/mL protein in 5 mM $MgCl_2$, 5 mM phosphonopyruvate and 10 mM Na^+Hepes (pH 7.0) were mixed with equal volumes of reservoir solution containing 12% polyethylene glycol 20,000 and 0.1 M MES (pH 6.0).

X-ray diffraction data for a SeMet-containing protein crystal were collected at IMCA-CAT 17-ID beam line at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Prior to data collection, crystals were transferred to mother

liquor containing 25% glycerol and flash-cooled in liquid propane cooled by liquid nitrogen. Multiple Anomalous Diffraction (MAD) data at 2.2Å resolution were collected from a single crystal at three different wavelengths (Se absorption edge inflection, peak, and high energy remote wavelength).

Table 3.1. *Data collection statistics*

Sample	**Wild-type**	**Se-Peak**	**Se-Inflection**	**Se-Remote**
Wavelength (Å)	1.0000	0.97944	0.97954	0.96485
Space Group	C2	C2		
Cell Dimensions (Å)	*a*=260.1 *b*= 83.8 *c*=104.9 *β*=112.1	*a*=258.7 *b*= 83.9 *c*=104.9 *β*=111.9		
Resolution (Å)	1.9 (1.97-1.9)	2.2 (2.28-2.20)	2.2 (2.28-2.20)	2.2 (2.28-2.20)
No. observations[a]	522013	526067	525654	526511
No. unique reflections	159433	190525	190709	190732
Completeness (%)	97(77.8)	91.8(94.7)	91.8(94.6)	91.6(94.3)
R_{merge}[b]	0.054(0.358)	0.042 (0.230)	0.042 (0.235)	0.042 (0.242)
Mean $I/\sigma(I)$	10.9(2.1)	39.1 (7.8)	21.3(4.1)	15.2(2.7)
Redundancy	3.3(2.1)	2.8(2.7)	2.8(2.7)	2.8(2.7)

[a] *Bijvot pairs are separated in the MAD experiment.*
[b] $R_{merge} = \sum_{hkl} [(\sum_j | I_j - < I > |) / \sum_j | I_j |]$. *Values in parentheses are for the highest resolution shell.*

Diffraction data for a wild-type PA4872 crystal were collected on the home facility using a Osmic Max-Flux monochromated X-rays generated by a Rigaku Micro Max 007 rotating anode equipped with a Rigaku AXIS IV^{++} detector. Data were processed with Crystal Clear/d*trek ((Pflugrath 1999), Rigaku/MSC, the Woodlands, TX). Data collection statistics are shown in Table 3.1.

3.3.5 Structure determination and refinement

Selenium sites were identified using the computer program SHELXD (Schneider and Sheldrick 2002). Initial phases were calculated using the program MLPHARE (Otwinowski 1991) in CCP4 (Cowtan 1994). Phase improvement by density modification followed (Cowtan 1994). The initial model was built automatically with the Solve/Resolve programs (Terwilliger and Berendzen 1999; Terwilliger 2003). Further building of side chains and model modification were done manually on a Silicon Graphics workstation using the graphics program "O" (Jones, Zou et al. 1991). After adding the side chains to one molecule, the asymmetric unit model was generated by applying the non-crystallographic symmetry operators.

Refinement was carried out using the data collected from the wild-type PA4872 crystal, which diffracted to higher resolution (1.9 Å) than the SeMet-containing protein crystal. CNS (Brunger, Adams et al. 1998) was used initially. Simulated annealing at 2500K was followed by alternating positional and individual temperature factor refinement cycles. As the refinement advanced, REFMAC (Cowtan 1994; Murshudov, Vagin et al. 1997) was used. Progress of the refinement was evaluated by the improvement in the quality of the sigma(a)-weighted $2F_{obs}$ - F_{calc} and F_{obs} - F_{calc} electron density maps (Read 1986), and the reduction of the conventional R-factor and R_{free}

values. Mg^{2+}, oxalate and water molecules were added gradually as the model improved. They were assigned in the F_{obs} - F_{calc} difference Fourier maps with a 3σ cutoff level for inclusion in the model. The refinement statistics are given in Table 3.2.

***Table 3.2**. Refinement statistics*	
Resolution (Å)	38.5 - 1.9
No. reflections refined	143540
No. reflections not refined (10%)	15883
$R_{work}/R_{overall}/ R_{free}$ [a]	0.195/0.202/0.257
No. Protein atoms	12767
No. Mg^{2+}-oxalate	6
No. water molecules	1728
RMSD from ideal geometry	
Bonds (Å)	0.019
Angles (°)	1.7
Mean B value $(Å)^2$	34

[a] *$R = \sum_{hkl} | |F_o| - |F_c| | / \sum_{hkl} |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{work} corresponds to the reflections used in refinement, R_{free} refers to randomly selected reflections omitted from the refinement, and $R_{overall}$ corresponds to all reflections.*

3.3.6 Modeling of the gating loop-closed conformation

A model of the active-site gating loop in a closed conformation was generated based on the C^{α} trace of the 2-methylisocitrate lyase loop. Side chains were added manually using the program "O". Energy minimization of the modeled loop was carried out in CNS, keeping the remainder of the molecule fixed.

3.3.7 Structure analysis

Superposition of structures was carried out within the program 'O' (Jones, Zou et al. 1991). The stereochemical quality of the structure was assessed with the program PROCHECK (Laskowski 1993). The programs PyMOL (DeLano 2002) and CCP4mg (Potterton, McNicholas et al. 2002; Potterton, McNicholas et al. 2004) were used for structure depictions. The program Voidoo was used to calculate the active site cavity (Kleywegt and Jones 1994).

3.3.8 Activity Assays for Substrate Screening

The catalyzed conversion of phosphonopyruvate to PEP was tested using the pyruvate kinase (PK)/ADP and lactate dehydrogenase (LDH)/NADH coupling system to convert PEP to lactate with the coupled conversion of NADH to NAD^+. The reaction solution (1 mL) contained 17 µM PA4872, 1 mM phosphonopyruvate, 5 mM $MgCl_2$, 20 units/mL PK, 20 units/mL LDH, 2 mM ADP, and 0.2 mM NADH in 50 mM K^+Hepes (pH 7.5 and 25°C). The absorbance of the solution was monitored at 340 nm ($\Delta\varepsilon$= 6.2 mM^{-1} cm^{-1}) over a 30 min time period.

The conversion of phosphonopyruvate to pyruvate was monitored using the LDH/NADH coupled assay. The reaction solution (1 mL) contained 17 µM PA4872, 1 mM phosphonopyruvate, 5 mM $MgCl_2$, 20 units/mL LDH, and 0.2 mM NADH in 50 mM K^+Hepes (pH 7.5 and 25°C). The absorbance of the solution was monitored at 340 nm ($\Delta\varepsilon$ = 6.2 mM^{-1} cm^{-1}) over a 30 min time period.

The lyase activity with the reactants (*R*)-malic acid, (*S*)-malic acid, (2*R*,3*S*)-isocitric acid, *threo*-(2*R*,3*S*:2*S*,3*R*)-2-methylisocitrate, 3-butylmalate, 2*R*-citramalate, 2*S*-citramalate, 2*R*-ethyl-3*S*-methylmalate and (2*R*,3*S*)-dimethylmalate was measured at

25°C using the LDH/NADH coupled assay. The reaction solutions (1 mL) initially contained 1 mM substrate, 17 µM PA4872, 5 mM $MgCl_2$, 20 units/mL LDH (except for 2*R*-ethyl-3*S*-methylmalate, for which 400 units/mL LDH was used), and 0.2 mM NADH in 50 mM K^+Hepes (pH 7.5 and 25°C). The absorbance of the solution was monitored at 340 nm ($\Delta\varepsilon$= 6.2 mM^{-1} cm^{-1}) over a 30 min time period. Control reactions not including PA4872 were run in parallel.

The conversion of carboxyPEP to phosphinopyruvate was tested using the malate dehydrogenase/NADH coupled assay. The reaction solution (1 mL) contained 17 µM PA4872, 1 mM carboxyPEP, 5 mM $MgCl_2$, 20 units/mL malate dehydrogenase, and 0.2 mM NADH in 50 mM K^+Hepes (pH 7.5 and 25°C). The absorbance of the solution was monitored at 340 nm ($\Delta\varepsilon$ = 6.2 mM^{-1} cm^{-1}) over a 30 min time period.

3.3.9 PA4872 catalyzed enolization of α-keto acids

The enolization rates of pyruvate, 2-keto-4-methyl pentanoate "α-ketoisocaproate", 2-keto butanoate "α-ketobutyrate", 2-keto pentanoate "α-ketovalerate", 2-keto-3-methyl butanoate "α-ketoisovalerate" and 2-keto-3-methyl pentanoate "α-keto-β-methyl-valerate" catalyzed by PA4872 were determined by monitoring the exchange of the C(3) hydrogen atom(s) in buffered D_2O using ^{1}H-NMR techniques. The solutions contained 100 mM Tris-d[11] (pH 7.5), 60 mM substrate, and 5 mM $MgCl_2$ in a total volume of 0.5 mL. Assays were initiated by the addition of PA4872 (1.3 µM for α-ketovalerate, 25 µM for pyruvate and α-ketobutyrate, 74 µM for α-ketoisovalerate and α-keto-β-methyl-valerate). ^{1}H-NMR spectra were recorded at 20 °C at 500 MHz with a Bruker Avance 500 NMR instrument spectrometer, using 10 mM 2-methyl-2-propan(ol-*d*) as the internal standard. The control reactions, which did not contain enzyme, were

carried out in parallel. The rates of α-ketovalerate enolization catalyzed by the H235A (8 μM) and H235Q (10 μM) PA4872 mutants were measured using the same protocol as were the rates of pyruvate enolization catalyzed by phosphonopyruvate hydrolase (5 μM), petal death protein (4 μM) and PEP mutase (5 μM).

The observed rate constant for catalyzed proton exchange (k_{enol}) was calculated from the ^{1}H-NMR spectra using equation 1 where A_t/A_0 is the ratio of the C(3) proton peak areas determined at time "t" and at t = 0, $[S_0]$ is the initial substrate concentration and [E] is the enzyme concentration.

$$k_{enol} = -(\ln(A_t/A_0))[S_0] / t[E] \quad (1)$$

3.3.10 ^{13}C-NMR analysis of PA4872 catalyzed decarboxylation of oxaloacetate

The ^{13}C-NMR spectra were recorded at 25 ^{o}C by using a Bruker Avance 500 NMR spectrometer. Methanol served as an internal standard (49 ppm). Following a 30 min incubation period of a solution initially containing 31 μM PA4872, 600 mM oxaloacetate, 5 mM $MgCl_2$ in deionized water at 25 ^{o}C (adjusted to pH 7.5 with aqueous NaOH), the PA4872 was removed by filtration using a centricon device. D_2O was added to the filtrate to a final concentration of 10% (v/v). The ^{13}C-NMR spectra of oxaloacetate, pyruvate, oxalate and bicarbonate in aqueous solution with 5 mM $MgCl_2$ (pH 7.5) were measured at 25^{o}C. Oxaloacetate signals appeared at 200.5 ppm (C(2)), 174.8 ppm (C(4)), 168.4 ppm (C(1)), 98.3 ppm (C(3)). The pyruvate signals appeared at 205.2 ppm (C(2)), 170.0 ppm (C(1)), 26.6 ppm (C(3)). The bicarbonate signal appeared at 161.2 ppm.

3.3.11 Oxaloacetate and 3-methyloxaloacetate decarboxylation assays

Direct Assay: the wild-type and mutant PA4872-catalyzed oxaloacetate decarboxylation reactions were monitored by using the continuous spectrophotometric assay for oxaloacetate consumption (decrease in solution absorbance at 255 nm; $\Delta\varepsilon$ = 1.1 mM^{-1}) (Lenz, Wunderwald et al. 1976). Reactions (0.25 mL) were carried out in quartz cells of 0.1 cm path length at 25 ^{o}C. Reaction solutions initially contained oxaloacetate, 5 mM $MgCl_2$ and 50 mM K^{+}Hepes (pH 7.5). Coupled assay: the PA4872 catalyzed decarboxylation reactions of oxaloacetate or 3-methyloxaloacetate were monitored at 340 nm ($\Delta\varepsilon$ = 6.2 mM^{-1} cm^{-1}) using a coupled assay for pyruvate or α-ketobutyrate formation. Reactions (1 mL volume) were carried out in quartz cells of 1 cm path length at 25 ^{o}C. Assay solutions contained oxaloacetate or 3-methyloxaloacetate, 10 U lactate dehydrogenase, 0.2 mM NADH, 5 mM $MgCl_2$ and 50 mM K^{+}Hepes (pH 7.5).

3.3.12 Steady-state kinetics analysis

The steady-state kinetic parameters (K_m and k_{cat}) were determined from the initial velocity data measured as a function of oxaloacetate concentration (varied from 0.5 to 5-fold K_m). The initial velocity data were fitted to Equation 2 with KinetAsystI:

$$V_0 = V_{max}[S]/(K_m +[S]) \quad (2)$$

where [S] is the substrate concentration, V_0 is the initial velocity, V_{max} is the maximum velocity, and K_m is the Michaelis-Menten constant. The k_{cat} value was calculated from V_{max} and the enzyme concentration using the equation $k_{cat} = V_{max}/[E]$, where [E] is the PA4872 subunit molar concentration in the reaction.

3.3.13 pH rate profile determination

The steady-state rate constants k_{cat} and k_{cat}/K_m were measured as function of reaction solution pH for PA4872-catalyzed oxaloacetate decarboxylation. The reactions were monitored using the direct assay (see above). Reaction solutions initially contained 5 mM $MgCl_2$, various concentrations of oxaloacetate (0.5-5 K_m) and PA4872 in the following buffers: (pH 5.0-5.5) 50 mM MES, (pH 5.5-8.0) 25 mM K^+Hepes and 25 mM MES, (pH 8.0-9.0) 25 mM K^+Hepes and 25 mM TAPS, (pH 9.0-9.5) 25 mM TAPS and 25 mM CAPSO.

3.3.14 ^{1}H-NMR analysis of PA4872 catalyzed 3-methyloxaloacetate decarboxylation

The reaction solution initially contained 5 mM $MgCl_2$, 10 mM 3-methyloxaloacetate, 6 nM PA4872, and 1 mM sodium acetate (internal standard) in 50 mM KH_2PO_4 (pH 7.0, 20 °C). ^{1}H-NMR spectra were recorded at 20°C with a Bruker Avance 500 NMR spectrometer. The decarboxylation reaction was monitored by measuring the disappearance of the C(3)CH_3 ^{1}H-NMR signal. The two control reactions, one which lacked PA4872, and the other both PA4872 and $MgCl_2$, were carried out in parallel.

The steady-state kinetic constant for catalyzed decarboxylation (k_{cat}) was calculated from the ^{1}H-NMR spectral data using equation 3, where A_t/A_0 is the ratio of the C(3) proton peak areas determined at time "t" and at t = 0, $[S_0]$ is the initial substrate concentration and [E] is the enzyme concentration.

$$k_{cat} = -(\ln(A_t/A_0))[S_0] / t[E] \quad (3)$$

3.3.15 Inhibition constant determination

The competitive inhibition constant K_i was determined for 2,2-difluoroxaloacetate, pyruvate, oxalate, α-ketovalerate, acetopyruvate, and phosphonopyruvate *vs.* oxaloacetate by fitting initial velocity data (obtained using the direct assay described above) to equation 4 with KinetAsystI

$$V_0=V_{max}[S]/[K_m(1 + [I]/K_i) + [S]] \quad (4)$$

Reaction solutions (0.25 mL) initially contained 0.8-8 mM oxaloacetate, 2×10^{-3} μM PA4872, 5 mM $MgCl_2$ and 3,3-difluoroxaloacetate (0, 0.8, 1.6 mM), pyruvate (0, 4, 8, 16 mM), oxalate (0, 0.1, 0.2 mM), α-ketovalerate (0, 8, 16 mM), acetopyruvate (0, 2, 4 mM), or phosphonopyruvate (0, 4, 8 mM) in 50 mM K^+Hepes (pH 7.5, 25 °C).

3.4 Results and discussion

A complete assignment of a biological function to an "unknown" member of an enzyme superfamily requires the identification of the type of chemical reaction that it catalyzes, the range of reactants that it can efficiently transform, and the metabolic context(s) of the reaction(s) in the organism. The active sites of the members of an enzyme superfamily have in common a conserved catalytic scaffold that positions core residues that function in stabilization of a common intermediate and/or transition state. Families within a superfamily diversify by acquiring active site residues that expand the range of chemical reactions. Accordingly, the core residues that stabilize the α-oxyanion intermediate shared by the PEP mutase/isocitrate lyase superfamily members are conserved in the PA4872 amino acid sequence. Key to the discovery of the structure of the intermediate stabilized by PA4872 is the determination of the spatial disposition of

the diversity residues within the catalytic site. Thus, our first step in the PA4872 function assignment was the determination of the PA4872 X-crystal structure.

3.4.1 Overall structure

The PA4872/Mg^{2+}-oxalate crystal structure, refined at 1.9 Å resolution, includes six molecules in the asymmetric unit. Four molecules denoted A-D form a tetramer (Figure 3.2A), the biological relevant assembly, and two molecules (E and F) form a dimer, which together with a crystallographic symmetry related dimer comprises another tetramer. The model includes residues 4-284 of molecule A except that no electron density is associated with the side chain of residue 128. Molecule B includes residues 4-128, and 130-285. Molecule C includes residues 4-255, and 259-284, with no electron density for the side chain of residue 128 is disordered. Molecule D includes residues 4-126 and 129-284. Molecule E includes residues 4-285 with no electron density for the side chain of residue 273. Molecule F includes residues 4-284 with no electron density for the side chains of residues 128 and 259. The model includes 1728 water molecules. The tetramers obey 222 symmetry and pack into a dimer of dimers. The root mean-square deviation (rmsd) between C^{α} atoms of the molecules in the asymmetric unit is on averages 0.2 Å. Each molecule contains Mg^{2+} and an oxalate. All backbone torsion angle values are within the allowed regions of the Ramachandran plot, and the bond lengths and bond angles are well within the range observed in crystal structures of small molecules (Table 3.2).

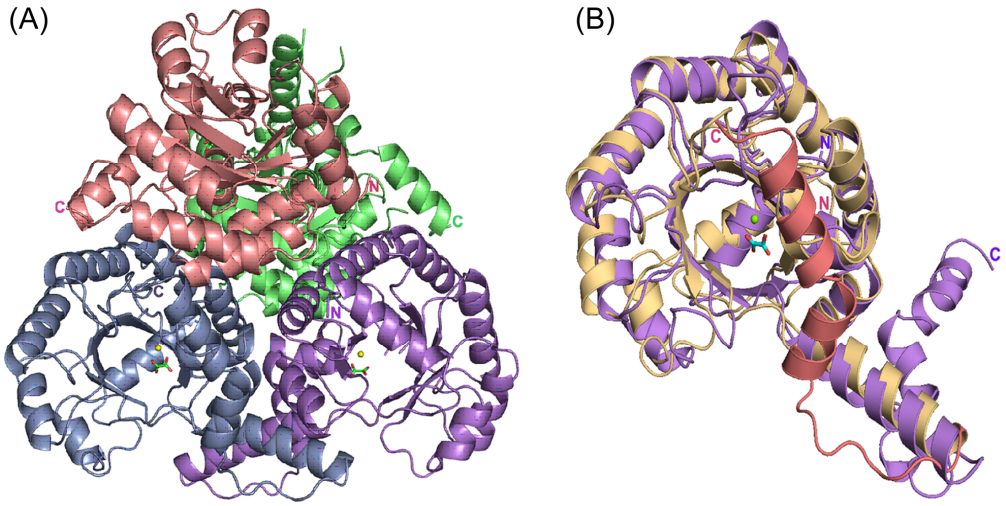


Figure 3.2 *– Overall fold of PA4872. (A) The tetramer structure highlighting each subunit in different color. Mg^{2+} is depicted as yellow sphere and oxalate as stick models with an atomic coloring scheme: Carbon - green, oxygen – red. (B) Superposition of the PA4872 (gold and salmon) and 2-methylisocitrate lyase (purple) monomers, illustrating the different paths of the C-terminal chains.*

PA4872 adopts the α/β barrel fold characteristic of the PEP mutase/isocitrate lyase superfamily, in which the eighth α-helix of a subunit is swapped with that of an adjacent subunit. The gating loop adopts an open conformation and the active site, containing Mg^{2+} and oxalate (Figure 3.2B), is solvent accessible. PA4872 differs from other superfamily members in the C-terminus following the swapped α-helix. Whereas the C-terminus of all other superfamily members traverses the active site of the partner subunit, in PA4872 the polypeptide chain turns back after the swapping of the helix and traverses its own active site (Figure 3.2B).

3.4.2 Active-site

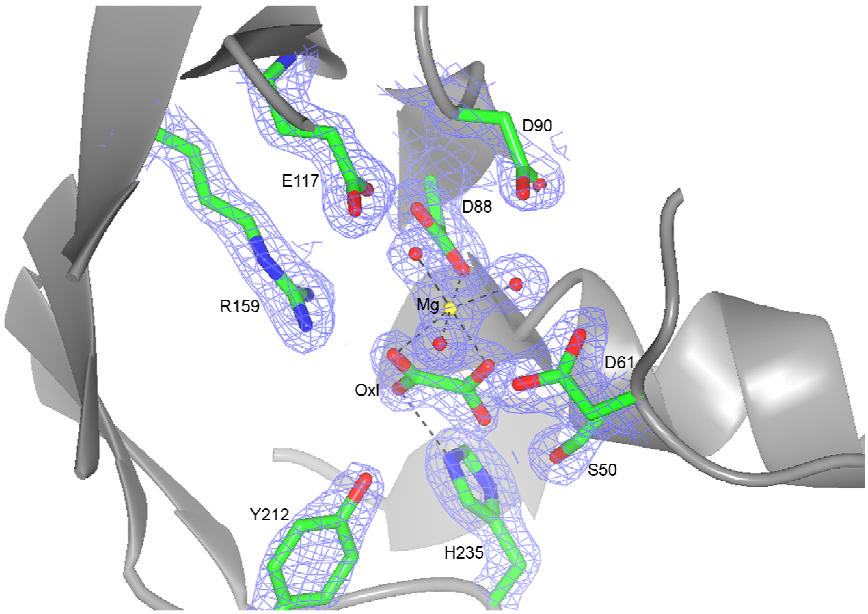
As with other α/β barrel enzymes (Branden 1980), the active site is located at the C-terminal end of the barrel's β-strands. Mg^{2+} and oxalate bind in a highly polar cavity (Figure 3.3). Mg^{2+} coordinates the C(1)O and C(2)O oxygen atoms of oxalate, the Asp88

carboxylate and 3 water molecules. The 3 water ligands engage in hydrogen bond interactions with the side chain carboxylic groups of Asp61, Asp90, Glu117, and Asp88. The oxalate C(1)OO- carboxylate interacts with the Ser50 hydroxyl group (2.5 Å), and with an oxyanion hole formed by the backbone amides of Gly49 and Ser50 (3.5Å and 3.1Å, respectively). Gly49 and Ser50 are located at the N-terminus of a short α-helix, suggesting that the helix dipole also contributes to substrate binding. The oxalate C(2)OO- engages in an ionic interaction with the guanidinium group of Arg159 (3.1 Å). No restraints were imposed on the planarity of the oxalate during the refinement. The six molecules in the asymmetric unit exhibit on average a 19º departure from planarity (referring to the C(1)-C(2) torsion angle). This appears to be a significant distortion but should be taken with caution because the oxalate crystallographic temperature factors are relatively high at 42 $Å^2$. The mode of oxalate interaction with the protein groups is the same as observed for the crystal structures of PEP mutase (Huang, Li et al. 1999) and phosphonopyruvate hydrolase (Chen, Han et al. 2006), and comparable to the mode of pyruvate interaction with protein groups observed for the crystal structure of 2-methylisocitrate lyase (Liu, Lu et al. 2005). The arginine, oxyanion hole and the Mg^{+2} binding residues comprise the core residues of the superfamily catalytic scaffold.

Two residues, located within the vicinity of the oxalate ligand, distinguish the PA4872 active site from those of other family members with known structure; Tyr212 positioned on the loop connecting β7 and α7, and His235 located on the loop connecting β8 and α8 (Figure 3.3). The His235 N^{ε} is located within hydrogen bonding distance (2.5-3.0 Å) of the oxalate C(1)O (the same C(1)O that engages in hydrogen bond interaction with the Ser50 hydroxyl group; the other C(1)O is coordinated to Mg^{2+}) and possibly

within hydrogen bonding distance of the oxalate C(2)O (3.2 -3.8 Å) (the other C(2)O is coordinated to Mg^{2+}).

(A)



(B)

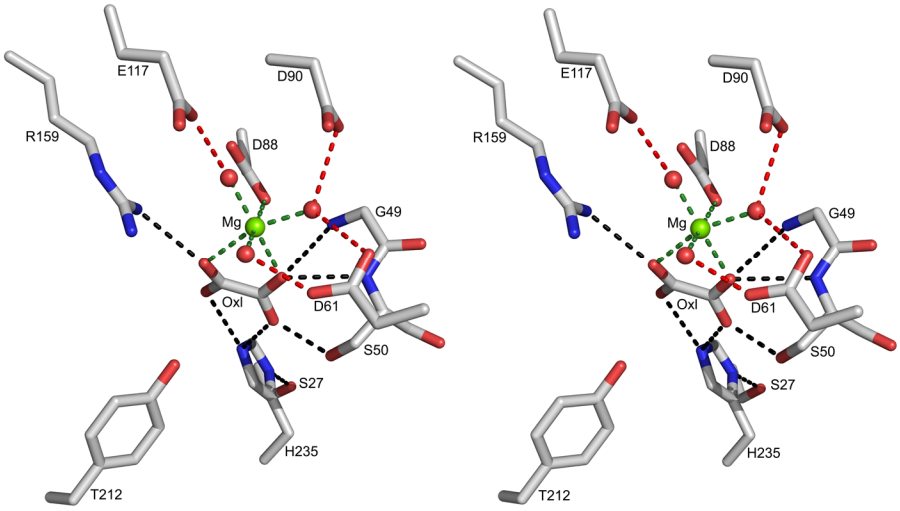


Figure 3.3 *- Active site of PA4871. (A) The electron density associated with key active site residues. The sigma(a)-weighted $2F_o - F_c$ map is shown contoured at 2σ level. Atomic colors are used: Carbon – green, oxygen – red, nitrogen – blue, Mg^{2+} - yellow. Oxalate is labeled Oxl (B) stereoscopic representation of the active site. Atomic colors are: Carbon – gray, oxygen – red, nitrogen – blue, Mg^{2+} - green.*

The H235 imidazole tautomer in which both the N^{ε} and N^{δ} atoms are protonated is evidenced by the hydrogen bond network associated with N^{δ}. Specifically, the His235 N^{δ} functions as a hydrogen bond donor and Ser27 hydroxyl (2.5 – 3.0 Å) as a hydrogen bond acceptor (Figure 3.3). In turn, Ser27 hydroxyl serves as a hydrogen bond donor to an internal water molecule that donates its hydrogen atom to the backbone carbonyl groups of Val28 and Ile46, and accepts a hydrogen bond from the backbone amide group of Gly48 (not shown). The conclusion that His235 is charged is consistent with the pH of the crystals (6.0). Within the PA4872 sequence family (Figure 3.4), His235 is stringently conserved and position 27 is occupied by a serine or a threonine. The hydrogen bond network suggests that the His235-Ser27 pair plays an important structural role in maintaining the fold integrity and active site architecture. However, site directed mutagenesis studies described below show that His235 does not play a major mechanistic role despite its close proximity to the oxalate and by inference to the substrate.

The Tyr212 hydroxyl group is positioned close to the oxalate ligand but not close enough for hydrogen bond interaction. Tyr212 is conserved among the majority of the PA4872 sequence family members (Figure 3.4). Although the gating loop is ordered in the PA4872 crystal structure, it does not close over the active site. Other examples of an "active site-open loop conformation" have been observed in X-ray structures of superfamily members that are known to require loop closure for catalytic turnover (Huang, Li et al. 1999; Liu, Lu et al. 2004; Liu, Lu et al. 2005; Chen, Han et al. 2006).

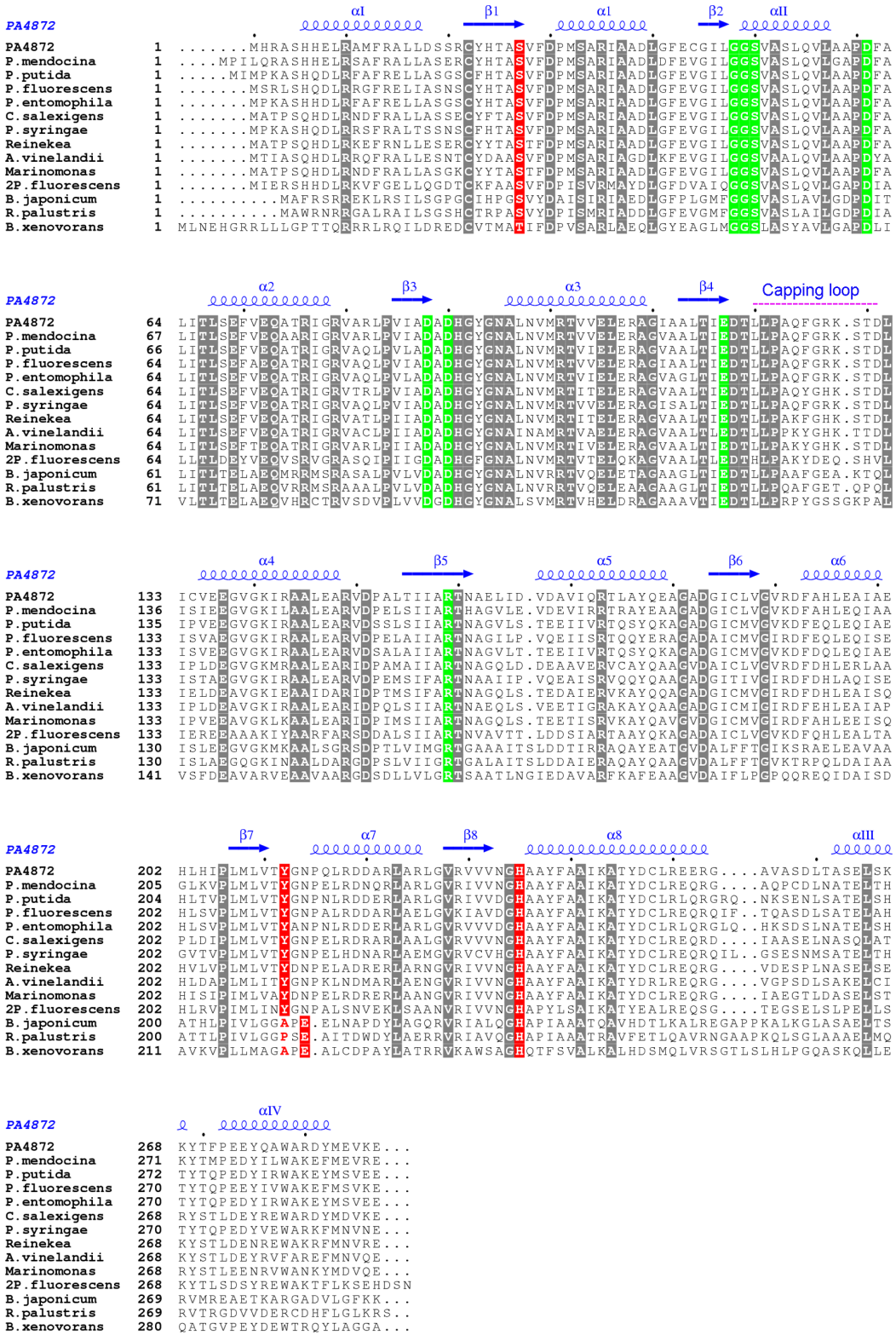


Figure 3.4 *– Structure based sequence alignment of PA4872 close relatives. Invariant active site residues that are conserved also in other PEP mutase/isocitrate lyase superfamily members are highlighted in blocked green. Other invariant residues are highlighted in blocked gray except for the active site His235, Tyr 212 (and the nearby glutamate of the 3-protein subgroup) and Ser27,*

which are highlighted in blocked red. Secondary structure motifs are shown above the sequences.

Loop closure is facilitated by the substrates or substrate analogs that engage in favorable electrostatic interaction with one or more loop residues. Because the oxalate ligand is not known to interact with the gating loop of other PEP mutase/isocitrate lyase superfamily members, the observation that the gating loop does not close over the PA4872 active site in the present structure is not substantial evidence that it does not do so during catalytic turnover. Yet, a key determinant of loop closure is the interaction between a loop lysine and a catalytic site glutamate residue (exemplified by Lys120 and Glu114 in PEP mutase (*7*)). PA4872 is unique in that the loop lysine residue is absent. For the superfamily members that require a loop closed conformation for catalytic turnover, the Glu-Lys interaction serves to pin the gating loop at one side of the active site.

To explore the potential contribution that the gating loop residues might make to the catalytic site we modeled the loop in the active site-closed conformation, using input from the structures of other superfamily members as a template (model not shown). Although we do not know the accuracy of this model, it does provide an indication that the loop does not contribute residues that might function in binding substrate *via* electrostatic interaction or participate in proton transfer (i.e. general acid/base catalysis). On the other hand, the model suggests that in the closed conformation, a conserved aromatic loop residue, Phe125 (replaced only by Tyr in the sequence family), may restrict solvent access to the active site.

A second structural feature of PA4872 that is unique within the superfamily is the α-helical C-terminus. Among the α/β-barrel fold structures, the PEP mutase/isocitrate

lyase superfamily is distinguished by the use of α/β-barrel 8th helix of one subunit to swap with that of the a neighboring subunit (Huang, Li et al. 1999). The α-helical segment that follows the 8th helix traverses the active site of the neighboring subunit and assists in its desolvation. The unique conformation of the PA4872 C-terminus, in which the main chain reverses direction after the α8 helix swapping, creates a channel leading to the active site adjacent the C-terminus of β8 (Figure 3.5A). None of the other PEP mutase/isocitrate lyase superfamily enzyme structures exhibit such a channel (see for comparison, the analogous surface representation of 2-methylisocitrate lyase in Figure 3.5B, which shows no such channel). The channel exposes the PA4872 active site to solvent from the side of the oxalate/pyruvate C(1) carboxylate group. Closure of the gating loop, as observed for the loop-closed model, does not close this channel, and therefore, we can assume that exposure of this region to solvent occurs during catalytic turnover. Solvent exposure occurs with the binary product complex of MOBH bound with ketopantoate (*11*), however this would not occur with the ternary substrate complex in which the N^5, N^{10}-methylene tetrahydrofolate "sits on top" of the active site, thereby shielding it from solvent while acting as the hydroxymethyl donor to the ensuing α-ketoisovalerate anion intermediate (Figure 3.1). The solvent accessibility of the PA4872 active site may by required for catalytic turnover, which among the superfamily members is unique.

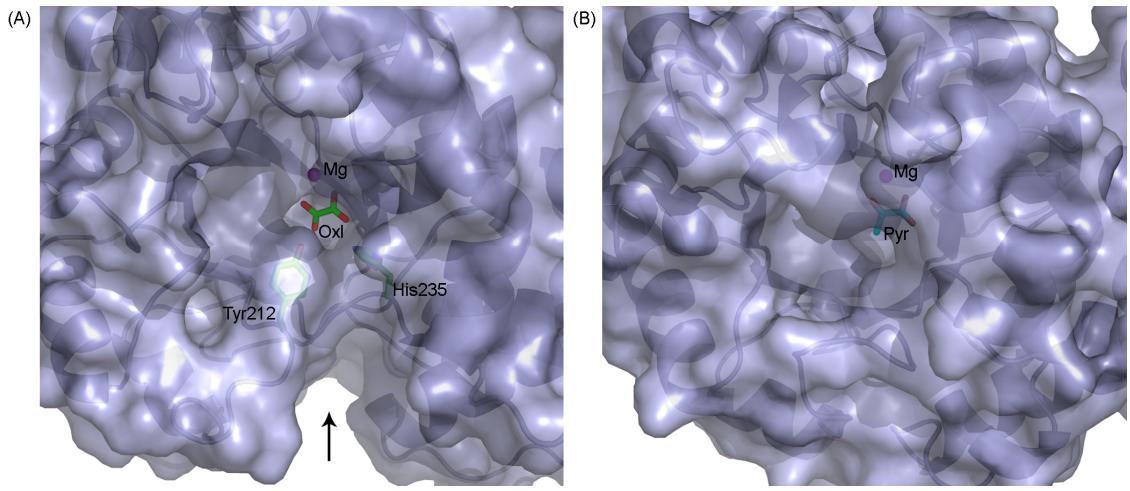


Figure 3.5 *– Molecular surfaces in the vicinity of active sites. (A) The channel leading to the active site of PA4872 (indicated by an arrow). Mg^{2+}-oxalate, His235 and Tyr212 are shown for reference. (B) The surface around the active site of 2-methylisocitrate lyase corresponding to the Mg^{2+}-pyruvate where the gating loop is in the open conformation (Liu, Lu et al. 2005).*

3.4.3 Assignment of PA4872 to the enolate-forming branch of the PEP mutase/isocitrate lyase superfamily

The reactions catalyzed by the PEP mutase/isocitrate lyase superfamily (Figure 3.1) fall into the two categories represented in Figure 3.6. The reactions catalyzed by PEP mutase, phosphonopyruvate hydrolase and MOBH proceed *via* a planar α-caboxy enolate intermediate (and/or transition state) whereas the reactions catalyzed by isocitrate lyase, 2-methylisocitrate lyase, petal death protein and oxaloacetate hydrolase proceed with the formation of a α-carboxy alkoxide and aci-carboxylate anion intermediate pair (and/or transition state). The conservation of the core residues known to stabilize the α-oxyanion intermediate, coupled with the absence of polar residues in the region of the active site that accommodates the aci-carboxylate anion intermediate formed by the C-C bond lyases (Britton, Langridge et al. 2000; Sharma, Sharma et al. 2000; Britton, Abeysinghe et al. 2001; Grimm, Evers et al. 2003; Simanshu, Satheshkumar et al. 2003; Liu, Lu et al. 2005; Teplyakov, Liu et al. 2005; Han, Joosten et al. 2007) suggested that

PA4872 might belong to the α-caboxy enolate intermediate forming branch of the superfamily. Activity tests designed to evaluate the ability of the PA4872 catalytic scaffold to stabilize an enolate anion intermediate confirmed this hypothesis.

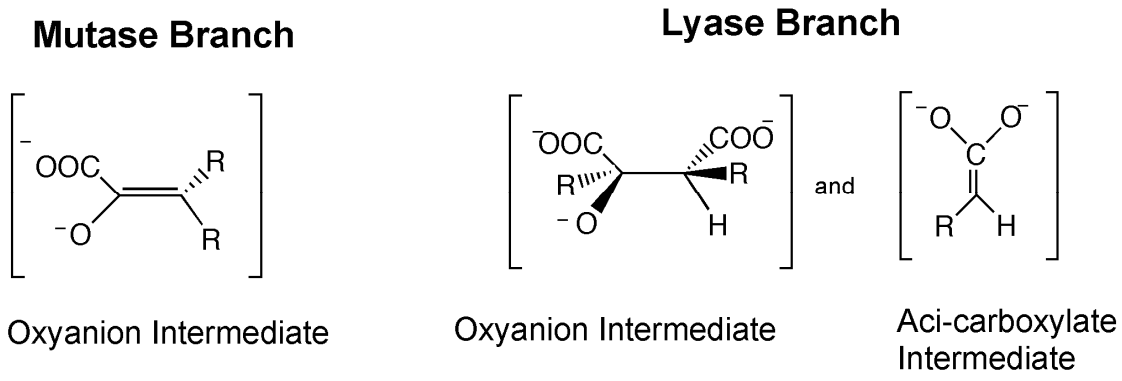


Figure 3.6 *–The type of intermediates formed by the mutase/transferase and lyase branches of the superfamily.*

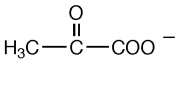
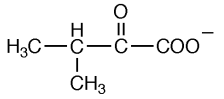
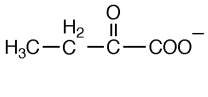
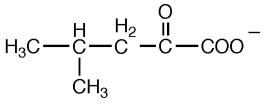
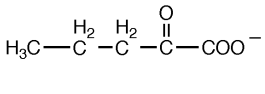
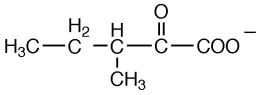
Blanchard and coworkers carried out solvent deuterium exchange reactions monitored by ^{1}H-NMR spectroscopy to demonstrate MOBH catalyzed enolization of α-ketoisovalerate (Sugantino, Zheng et al. 2003). Although PA4872 lacks the MOBH general base (Glu181 in the *E. coli* enzyme (von Delft, Inoue et al. 2003)) that abstracts the C(3)H of the α-ketoisovalerate in the enolization partial reaction (Val188 occupies this position in PA4872), PA4872 does conserve the electropositive residues and the Mg^{2+} cofactor that function in stabilization of the α-carboxy enolate anion. Using the same ^{1}H-NMR based approach, the rate constants for the enolization (k_{enol}) of pyruvate catalyzed by phosphonopyruvate hydrolase and PEP mutase were measured. Like PA4872 these enzymes conserve the electropositive residues and the Mg^{2+} cofactor that stabilize the pyruvate enolate intermediate but lack an enzyme residue to function as a general base in the enolization of pyruvate. The k_{enol} measured for phosphonopyruvate hydrolase is 12 min^{-1} and that for PEP mutase is 4 min^{-1}. These rate constants are

significantly smaller than the k_{cat} values measured for the catalyzed reactions shown in Figure 3.1 (105 s^{-1} (Chen, Han et al. 2006) and 34 s^{-1} (Kim, Kim et al. 1998), respectively) and thus, the turnover rates greatly exceed the enolization rates. In contrast, the reported k_{cat} for the *Mycobacterium tuberculosis* MOBH is 47 min^{-1} and the rate constant for the enolization of α-ketoisovalerate (k_{enol}) is 752 min^{-1} (Sugantino, Zheng et al. 2003), which greatly exceeds the turnover rate. The difference in kinetic behavior is consistent with the fact that MOBH catalyzes the enolization of α-ketoisovalerate as the first partial reaction of the overall reaction (Figure 3.1) and that MOBH possesses a general base (Glu181) to assist in the catalysis of the α-ketoisovalerate enolization. Phosphonopyruvate hydrolase and PEP mutase, on the other hand, do not catalyze pyruvate enolization as a partial reaction of the overall reaction, nor do they possesses a general base that might facilitate pyruvate enolization. We assume that a water molecule functions in base catalysis of the pyruvate enolization. The rate enhancements observed with phosphonopyruvate hydrolase and PEP mutase are attributed to the delocalization of excess charge from the pyruvate enolate onto the Mg^{+2} cofactor and the electropositive active site residues.

The k_{enol} determined for the petal death protein is 0.6 min^{-1} whereas isocitrate lyase and 2-methylisocitrate lyase do not catalyze exchange at a detectable level. These three enzymes do not form an enolate anion intermediate (Figure 3.1). Based on these results we conclude that the 2 branches of the superfamily are distinguished by the ability *vs*. inability to catalyze the enolization of α-ketoacids. In this context we interpret the k_{enol} = 24 min^{-1} measured for PA4872 catalyzed pyruvate enolization (Table 3.3) as evidence

that it is a member of the α-carboxy enolate anion intermediate forming branch of the superfamily.

Table 3.3. *The rate constant for enolization (k_{enol}) determined from the rate of PA4872-catalyzed deuterium exchange from D_2O solvent with the C(3)H of 2-ketoacids*
[a] The rate constant for the PA4872 H235Q mutant.
[b] The rate constant for the PA4872 H235A mutant.

Linear 2-ketoacids	k_{enol} (min^{-1})	Branched 2-ketoacids	k_{enol} (min^{-1})
 pyruvate	24	 α-ketoisovalerate	0.2
 α-ketobutyrate	56	 α–ketoisocaproate	5
 α-ketovalerate	270 82 [a] 77 [b]	 α-keto-β-methyl-valerate	< 0.2

To gain information about the structure of the α-carboxy enolate anion intermediate formed during PA4872 catalytic turnover of its physiological substrate, the k_{enol} for a variety of C(3) substituted pyruvate compounds were measured (Table 3). Whereas only α-ketoisovalerate is substrate for the overall reaction catalyzed by MOBH (Sugantino, Zheng et al. 2003), this enzyme catalyzes the enolization of pyruvate and C(3) substituted pyruvate with the following efficiencies: for the native substrate α-ketoisovalerate k_{enol} = 752 min^{-1}; for α-keto-□-methyl-valerate k_{enol} = 482 min^{-1}; for α-ketobutyrate k_{enol} = 354 min^{-1}; for α-ketovalerate k_{enol} = 229 min^{-1}; for pyruvate k_{enol} = 129 min^{-1}; and for α-ketoisocaproate k_{enol} = 19 min^{-1}. In contrast, PA4872 is only significantly active with the linear α-ketoacids for which the order of reactivity is α-

ketovalerate > α-ketobutyrate > pyruvate (Table 3.3). This result suggested that the native substrate is a small linear α-ketoacid.

3.4.4 PA4872 chemical function determination

The PA4872 active site provides space in a hydrophobic pocket lined by Met208, Val210 and Val232, for a small C(3) substituent such as a methyl or ethyl group. Adjacent to the Tyr212, there is also space for a second C(3) substituent oriented towards the active site exit at the C-terminus edge of the α/β barrel, albeit with slight adjustment of the Tyr212 side chain conformation. This suggested the possibility that PA4872 catalyzes a β-elimination reaction in an α-ketoacid substrate. The obvious candidate for such a reaction is oxaloacetate. The PA4872 active site, modeled with oxaloacetate is shown in Figure 3.7 along with models built with 3-methyloxaloacetate and 3-ethyloxaloacetate as the ligands. All three models conserve the electrostatic interactions with the core residues that stabilize the pyruvate enolate while the carboxylate group is confined to a position above the plane of the developing pyruvate enolate intermediate, in accord with the accepted geometry for decarboxylation in a β-ketoacid (O'Leary 1992).

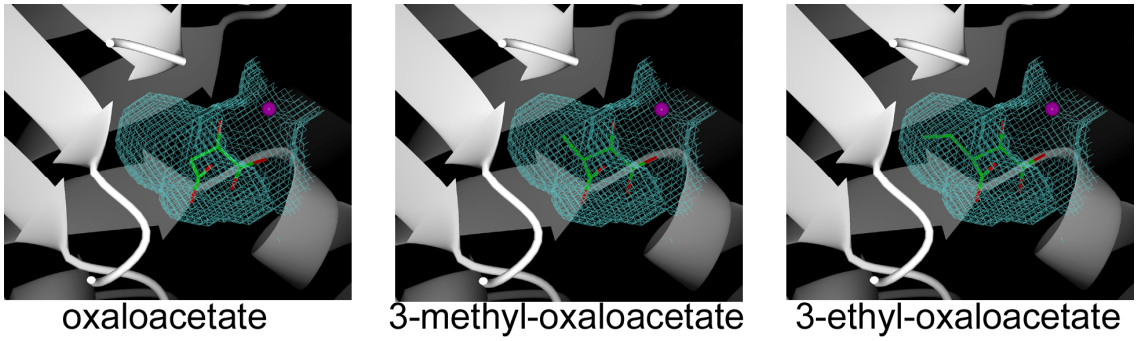


Figure 3.7 *– The active site cavity (shown as a cage) with docked potential substrates shown as stick models with the atomic colors green – carbon, red – oxygen. The Mg^{2+} is depicted as a magenta sphere.*

The addition of PA4872 to a buffered solution of oxaloacetate and Mg^{2+} was accompanied by rapid bubble formation (*viz.* CO_2). ^{13}C-NMR analysis of the solution confirmed that the oxaloacetate had been quantitatively converted to pyruvate and bicarbonate. The steady state rate constants for this reaction were measured using two different assay methods. The first monitors the disappearance of oxaloacetate (absorbance decrease at 255 nm) and the second monitors the formation of pyruvate *via* coupled reaction with lactate dehydrogenase and NADH (absorbance decrease at 340 nm). The results are presented in Table 3.4.

Table 3.4*. Steady-state kinetic constants for PA4872 catalyzed decarboxylation of oxaloacetate and 3-methyloxaloacetate*

Substrate	**Assay**	k_{cat} **(s^{-1})**	K_m **(mM)**	k_{cat}/K_m **($M^{-1}s^{-1}$)**
oxaloacetate	direct	7500 ± 200	2.2 ± 0.1	3×10^6
oxaloacetate	coupled	7200 ± 400	1.3 ± 0.2	6×10^6
3-methyl-oxaloacetate	coupled	250 ± 10	0.63 ± 0.08	4×10^5
3-methyl-oxaloacetate [a]	^{1}H-NMR	800	--------	--------

[a] *The reaction mixture in this assay initially contained 10 mM 3-methyloxaloacetate, 5 mM $MgCl_2$, 6 nM PA4872, 50 mM KH_2PO_4 (pH 7.0, 20 ℃), and 1 mM sodium acetate (as the internal standard). The ^{1}H-NMR of 3-methyloxaloacetate measured under the reaction conditions defines the relative ratio of enol:ketone:gem diol as 1.7: 0.8: 0.8. The rate of decarboxylation of 3-methyloxaloacetate in buffer is 0.052 mM/min, the rate in $MgCl_2$-containing buffer is 0.078 mM/min whereas the rate measured in the presence of the 6 nM PA4872 is 0.38 mM/min.*

Owing to the fact that oxaloacetate is a slow substrate for lactate dehydrogenase (hence a background rate must be subtracted) the direct assay is likely to be the more accurate measurement. The k_{cat} = 7500 s^{-1} is 2-3 orders of magnitude larger than the k_{cat} of PEP mutase (k_{cat} = 34 s^{-1} for the enzyme from *Mytilus edulis (Kim, Kim et al. 1998)*), phosphonopyruvate hydrolase (k_{cat} = 105 s^{-1} for the enzyme from *Variovorax* sp. Pal2 (Chen, Han et al. 2006)), oxaloacetate hydrolase (k_{cat} = 10 s^{-1} for the enzyme from

Botrytis cinerea (Han, Joosten et al. 2007)), isocitrate lyase (k_{cat} = 100 s^{-1} for the enzyme from *E. coli* (Liu, Lu et al. 2005)), 2-methylisocitrate lyase (k_{cat} = 19 s^{-1} for the enzyme from *E. coli* (Liu, Lu et al. 2005)), and petal death protein (k_{cat} = 8 s^{-1} for the best substrate (Lu, Feng et al. 2005)) in reaction with their physiological substrates. The k_{cat} is 4 orders of magnitude larger than that of the MOBH (k_{cat}= 0.8 s^{-1}).

Knowledge of the k_{cat} for PA4872 catalysis of its physiological reaction allows us to contrast the efficiency by which the pyruvate enolate is formed by decarboxylation of oxaloacetate to the efficiency at which the enzyme catalyzes the enolization of bound pyruvate; at k_{cat} = 7500 s^{-1} *vs.* k_{enol} = 0.4 s^{-1}, the trend is the same as PEP mutase and phosphonopyruvate hydrolase, and the reverse of MOBH. Although the His235 N$^{\square}$ is located near the C(3) of the modeled oxaloacetate it does not make a significant contribution to catalysis of α-ketoacid enolization. Specifically, k_{enol} values measured for H235A and H235Q PA4872 with the best enolization substrate (α-ketovalerate) are only 3-fold smaller than that of wild-type PA4872 (Table 3.3).

The models of PA4872 bound with 3-methyl- and 3-ethyloxaloacetate (Figure 3.7) suggested that these compounds might also be good substrates for PA4872 decarboxylation. 3-methyloxaloacetate is a metabolite formed from L-threo-3-methylaspartate within the C-5 branched dibasic acid metabolic pathway (http://www.genome.ad.jp/kegg/pathway.html; (Kanehisa, Goto et al. 2006)) and thus a potential physiological substrate for PA4872. 3-Methyloxaloacetate was prepared by chemical synthesis (see Supporting Information) and the rate of PA4872 decarboxylation was monitored by ^{1}H-NMR to define a turnover rate of 790 s^{-1} at 10 mM 3-methyloxaloacetate (Table 3.4). The steady-state kinetic constants were measured using

the LDH/NADH coupled assay for formation of α-ketobutyrate defining k_{cat} = 250 ± 10 s^{-1}, K_m = 0.63 ± 0.08 mM and k_{cat}/K_m = 4 × 10^5 M^{-1} s^{-1}. Although the k_{cat}/K_m value is an order of magnitude smaller than that measured for oxaloacetate, it is still large enough to support the possible physiological relevance of this reaction. Whereas, the C(3)CH_3 substituent increased the rate of the enolization of the α-ketoacid 2-fold (α-ketovalerate k_{enol} = 56 min^{-1} *vs*. pyruvate k_{enol} = 24 min^{-1}) it decreased the rate of decarboxylation in the corresponding diacid 3-methyloxaloacetate relative to oxaloacetate by ~20-fold. Thus, the PA4872 catalyzed decarboxylation reaction is hindered by the C(3)CH_3 substituent and based on this observation we hypothesize that the PA4872 evolved to function in the decarboxylation of the main stream metabolite oxaloacetate rather than the decarboxylation of the more obscure metabolite 3-methyloxaloactate.

3.4.5 PA4872 catalytic proficiency

The rate of oxaloacetate decarboxylation in aqueous solution was investigated earlier using 3-methyloxaloacetate in place oxaloacetate so that the reaction can be monitored by ^{1}H-NMR. The rate measured at pH 4.6 and 31^{o}C was 7.8 x10^{-5} s^{-1} (Kubala and Martell 1981). The rate does not increase at higher pH nor does the presence of Mg^{+2} has a large impact on the reaction (~2-fold increase). Thus, the approximate value for $k_{catalyzed}/k_{solution}$ is 10^8.

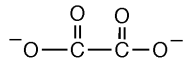
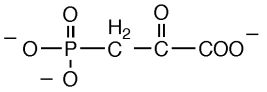
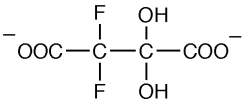
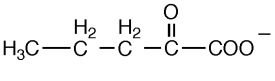
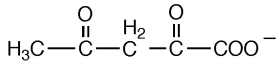
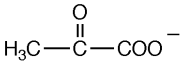
The steady state kinetic constants measured for decarboxylation of oxaloacetate catalyzed by PEP mutase are k_{cat} =0.269 ± 0.004 s^{-1} and K_m = 0.22 ± 0.01 mM (k_{cat}/K_m = 1.2 × 10^3 M^{-1} s^{-1}). The values measured for phosphonopyruvate hydrolase are k_{cat} = 1.04 ± 0.02 s^{-1} and K_m = 1.6 ± 0.1 mM (k_{cat}/K_m = 6.5 × 10^2 M^{-1} s^{-1}). Our "control enzyme" 2-methylisocitrate lyase did not catalyze oxaloacetate decarboxylation at a detectable rate

(k_{cat} < 1 x 10^{-3} s^{-1}). Thus, the ability of the members of the pyruvate enolate intermediate forming branch of the PEP mutase/isocitrate lyase superfamily to stabilize the pyruvate enolate anion (Figure 3.6) is sufficient to confer a weak decarboxylase activity on them. The proficiency of the PA4872 ($k_{catalyzed}/k_{solution}$ ~ 10^8) is substantially greater than that of PEP mutase ($k_{catalyzed}/k_{solution}$ ~ 10^3) and phosphonopyruvate hydrolase ($k_{catalyzed}/k_{solution}$ ~ 10^4). Hence, PA4872 diverged in function through optimization of the an active site environment to facilitate decarboxylation.

3.4.6 PA4872 Substrate Recognition

The binding constants for substrate and intermediate analogs were determined by measuring the competitive (*vs.* oxaloacetate) inhibition constant (K_i) for these compounds. The pyruvate enolate analog used in the X-ray structure determination, oxalate, was found to have a K_i = 43 μM (Table 3.5). This value is comparable to the K_i = 25 μM measured for PEP mutase (Liu, Lu et al. 2002). In contrast, phosphonopyruvate, which binds very tightly to PEP mutase (K_m = 1 μM and K_i for sulfopyruvate = 40 μM) (Kim, Kim et al. 1998; Liu, Lu et al. 2002), binds very poorly to PA4872 (K_i = 3 mM). Thus, the two enzymes share tight binding to the pyruvate enolate analog but diverge in substrate recognition. Oxalate is also the product of oxaloacetate hydrolase and petal death protein catalyzed oxaloacetate cleavage. The K_i values measured for these two enzymes are 19 □M (Han, Joosten et al. 2007) and 4 □M (Lu, Feng et al. 2005), respectively.

Table 3.5. *The competitive inhibition constant determined for PA4872 catalyzed oxaloacetate decarboxylation.* [a] *The inhibition constant was measured for the PA4872 Y212F mutant rather than the wild type PA4872.*

Inhibitor	K_i (mM)	Inhibitor	K_i (mM)
 oxalate	0.043 ± 0.002	 phosphonopyruvate	3.0 ± 2
 3,3-difluoroxaloacetate	0.45 ± 0.03	 α-ketovalerate	6.7 ± 0.3 3.2 ± 0.3[a]
 acetopyruvate	1.09 ± 0.08 0.56 ± 0.04[a]	 pyruvate	7.2 ± 0.4

3,3-Difluoroxaloacetate, which differs from oxaloacetate by replacement of the C(3) hydrogens with fluorine atoms, is not a substrate for PA4872 decarboxylation. Nor is it a substrate for oxaloacetate hydrolase catalyzed hydrolytic cleavage (Han, Joosten et al. 2007). Whereas oxaloacetate exists predominantly in the C(2) keto form (Lu, Feng et al. 2005), the 3,3-difluoroxaloacetate exists predominantly in the C(2) gem diol (hydrate) form (Han, Joosten et al. 2007). Oxaloacetate hydrolase catalyzes C(2)-C(3) bond cleavage in the oxaloacetate gem diol. The tight binding of 3,3-difluoroxaloacetate (K_i = 68 nM, (Han, Joosten et al. 2007)) suggests that the enzyme active site is specialized in recruiting or generating the gem diol form of oxaloacetate. It is not clear why the fluorine atoms at C(3) impair the C(2)-C(3) cleavage step. PA4872 binds 3,3-difluoroxaloacetate 7500-fold less tightly than does oxaloacetate hydrolase at K_i = 450 □M and is unable to decarboxylate this substrate analog. The comparatively higher stability of the 3,3-difluoroxaloacetate to decarboxylation in solution and in the presence of PA4872 may be

due to the predominance of the gem diol and/or another inherent property of the keto form. The K_i values of α-ketovalerate (6.7 mM) and pyruvate (7.2 mM) were measured to determine whether the C(3) ethyl substituent enhances binding affinity, which it does not (Table 3.5). Lastly, we synthesized acetopyruvate (see Supporting Information) to test as an oxaloacetate analog. The K_i = 1 mM indicates a ~7-fold tighter binding than observed for pyruvate.

3.4.7 PA4872 Catalytic Mechanism

Shown in Figure 3.8 are the key interactions of oxaloacetate modeled in the PA4872 active site. These are analogous to the set of interactions of the phosphonopyruvate substrate in the actives sites of PEP mutase and phosphonopyruvate hydrolase, which share with PA4827 the capacity to catalyze oxaloacetate decarboxylation (albeit at a much reduced rate). The modest oxaloacetate decarboxylase activity of PEP mutase and phosphonopyruvate hydrolase is evidence that these interactions are applicable to oxaloacetate binding and decarboxylation and that the difference in catalytic efficiency originates in difference in the active site environments. Not shown in Figure 3.8 are the two polar residues of the PA4872 active site with the greatest potential to facilitate the decarboxylation are His235 and Tyr212 (Figure 3.3), conserved residues in the PA4872 sequence family (Figure 3.4) that are absent in PEP mutase and phosphonopyruvate hydrolase. At first glance one might envision hydrogen bond interaction between the Tyr212 hydroxyl group and the oxaloacetate C(3)COO- and a role for the His235 in general acid catalysis. Further analysis, however did not support either catalytic role because replacements of these two residues did not impact enzyme efficiency significantly (Table 3.6).

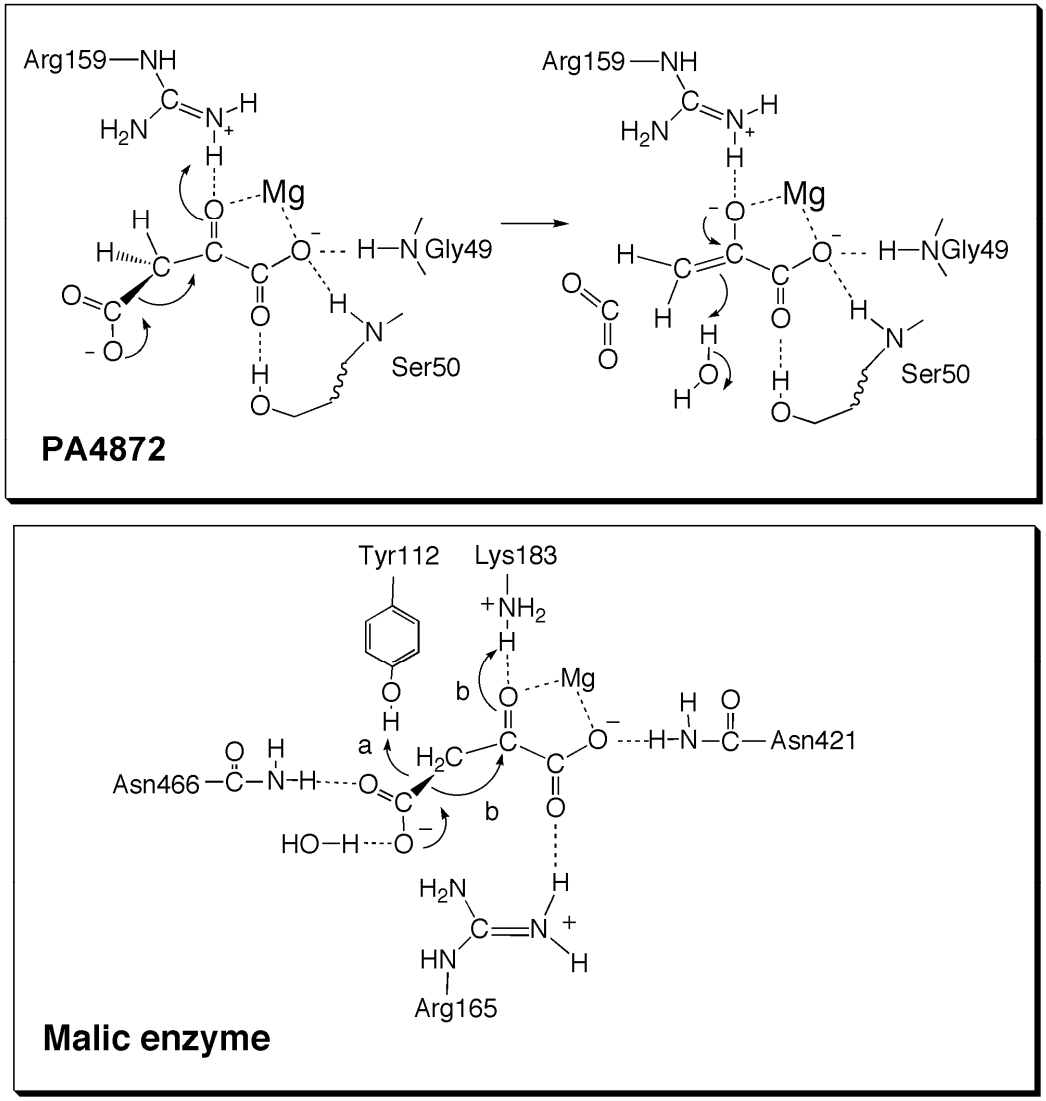


Figure 3.8 *– Proposed catalytic mechanism of PA4872 and comparison with the catalytic machinery of the malic enzyme oxaloacetate decarboxylase.*

Perhaps noteworthy is the 25-fold reduction in the K_m value of the Y212F mutant, which suggests that the Tyr212 hydroxyl group impairs substrate binding yet at the same time it facilitates catalysis (the k_{cat} of Y212F PA4872 is reduced by 23-fold). This conjecture is supported by the observation that the K_i values measured for acetopyruvate (aceto group in place of the oxaloacetate C(3) carboxylate group) and α-ketovalarate

(ethyl group in place of the oxaloacetate C(3) carboxylate group) are 2-fold smaller than those measured for the wild-type enzyme (Table 3.5).

***Table 3.6**. Steady-state kinetic constants for mutant PA4872 catalyzed decarboxylation of oxaloacetate.*[a]

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
WT	7500 ± 200	2200 ± 100	3.4×10^6
Y212F	323 ± 6	87 ± 4	3.7×10^6
H235A	510 ± 7	1020 ± 50	5.0×10^5
H235Q	2580 ± 90	2500 ± 200	1.0×10^6

[a] *0.25 mL reaction mixture contained 50 mM K^+Hepes (pH 7.5, 25 °C), 5 mM $MgCl_2$, 0.5-10 K_m of oxaloacetate. Reactions were monitored at 255 nm.*

Relevant to the catalytic mechanism, the pH rate profile data (Figure 3.9) shows that there is no change in the k_{cat} value over the pH 5-10 range and there is not sufficient variation in the k_{cat}/K_m to assign to the ionization of an essential residue. This rules out that a protonated His235 serves as the general acid in the decarboxylation reaction. Otherwise we would anticipate a drop in catalytic efficiency above pH 6.5. Moreover, the k_{cat} and k_{cat}/K_m values determined for H235A and H235A catalysis of oxaloacetate decarboxylation are less than an order of magnitude smaller than the k_{cat} and k_{cat}/K_m value measured for wild-type PA4872. The stringent conservation of the histidine residue among the PA4872 sequence family members is attributed to its structural role (described in a previous section) rather than to a catalytic role. Because Tyr212 and His235 do not appear to play a direct role in the catalytic mechanism, they are not included in Figure 3.8.

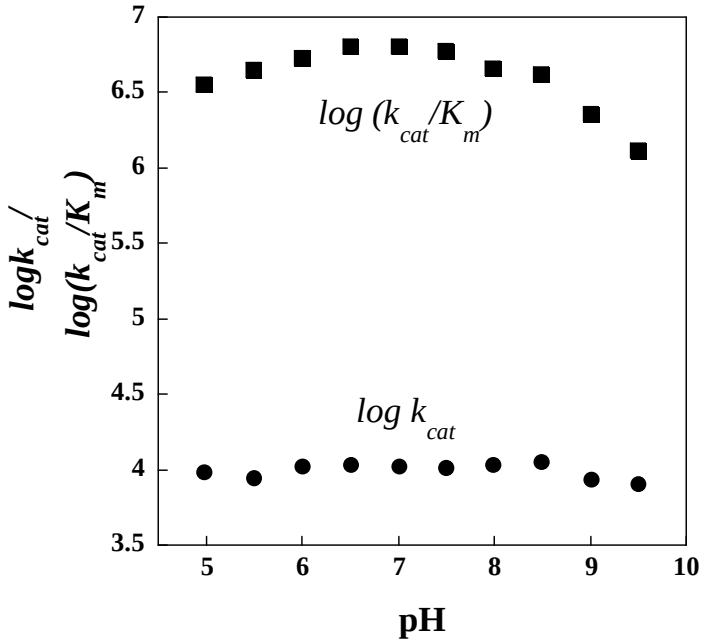


Figure 3.9 *– pH profile of PA4872. The buffers used in various pH ranges are listed in the Methods section*

The origin of the greater rate enhancement observed with PA4872 *vs*. the PEP mutase and phosphonopyruvate hydrolase is currently unknown but may arise from two unique structural features. The first is a sterically restricted hydrophobic region of the active site which is able to confine the C(3)COO- in an orientation optimal for the dissociation of CO_2 (Figure 3.7). If the gating loop closes, the desolvation of the C(3)COO- into a nonpolar region of the active site is also expected to contribute to ground state destabilization that is relieved as the charge is transferred to the C(3) upon CO_2 dissociation. The second important feature of the PA4872 active site is the solvent access to C(3) provided by the channel (Figure 3.6), which is unique to PA4872. A water molecule that approaches C(3) from that direction may serve as the acid that protonates the pyruvate enolate.

There are several other enzyme superfamilies from which an oxaloacetate decarboxylase has emerged. Pyruvate kinase possess an intrinsic oxaloacetate decarboxylase activity owing to its ability to stabilize a pyruvate enolate intermediate, but oxaloacetate is not considered its physiological substrate (Creighton and Rose 1976; Jursinic and Robinson 1978). The malic enzyme catalyzes oxaloacetate decarboxylation as the second partial reaction in the conversion of malate to CO_2 plus pyruvate with the reduction of NAD to NADH (Park, Harris et al. 1986). CitM within the malic enzyme family has evolved to function in a specialized citrate degradation pathway in a small group of Gram- positive bacteria (Sender, Martin et al. 2004). The CitM cannot convert malate to oxaloacetate but it does decarboxylate oxaloacetate at an efficiency of k_{cat} = 20 s^{-1} and K_m = 0.5 mM using the same catalytic residues of the malic enzyme. The catalytic residues of CitM and the malic enzyme (Chang and Tong 2003) are depicted in Figure 3.8 along with two possible pathways for the catalyzed decarboxylation. As with PA4872, a Mg^{+2} cofactor is used along with several electropositive active residues to bind the α-keto acid moiety. There are however distinct differences in catalytic mechanism. Firstly, the carboxylate is oriented for the dissociation by binding to an asparagine side chain and an internal water molecule. Secondly, a tyrosine residue is positioned close to the C(3) so that it might function as the acid catalyst (pathway a). Alternatively, the Lys residue that engages in hydrogen bond formation with the substrate C(2)=O may transfer its proton as the reaction proceeds (pathway b). Despite its seemingly more sophisticated catalytic site, the catalytic efficiency of CitM is significantly lower than that of PA4872.

A third known "dedicated" oxaloacetate decarboxylase is the oxaloacetate decarboxylase Na^+ pump present in a wide range of bacterial species. This membrane protein system catalyzes the transfer of the C(3) carboxyl group of oxaloacetate to a biotin prosthetic group and the decarboxylation of the carboxybiotin, coupled to Na^+ translocation. The recent crystal structure of the enzyme shows that the Zn^{2+}-dependent oxaloacetate decarboxylase domain adopts an α/β barrel fold, but otherwise the catalytic mechanism is entirely different than that of PA4872 (Studer, Dahinden et al. 2007); a catalytic lysine residue serves as the acceptor of the CO_2 from oxaloacetate and transfers it to biotin, which induces the sodium translocation.

A fourth dedicated oxaloacetate decarboxylase from Pseudomonas species has been characterized. This large (~600 amino acids) 3-domain decarboxylase is found in a wide range of Gram-negative bacteria, including the same species of bacteria that produce the PA4872 family members. Although the structure of this enzyme has not been determined, the stereochemistry of the protonation with inversion at C(3) has been reported (Piccirilli, Rozzell et al. 1987) as well as the results from an in depth examination of transition state structure, metal cofactor activation and pH effects (Waldrop, Braxton et al. 1994).

3.4.8 PA4872-like species distribution and biological function

PA4872 orthologs are found in all deposited Pseudomonad genomes (genus of γ-proteobacteria) and in only a few other Gram-negative bacteria. The most distinguishing sequence features in this group of enzymes is the invariant His235-Ser27 (or Thr) pair, the lack of the conserved lysine that regulates the closure of the gating active site loop, and the very different sequence of the gating loop compared with other enzymes in the

PEP mutase/isocitrate lyase superfamily (Figure 3.4). Two of the Pseudomonad species each have a strain that contain two copies of the gene (paralogs), well separated within the genome. One paralog shares high sequence identity with PA4872 (>70%) while the other shares 50-60% identity. The identity level of both paralogs within a genome is sufficiently high to indicate the same chemical function although their substrate specificity may vary. Within the Pseudomonad group the sequence identity between PA4872 and the closest homologues is 70-99 %. Outside of the Pseudomonad group it is in the 70-80% range for *Chromohalobacter salexigens, Reinekea* sp. MED297 and *Marinomonas* sp. MWYL1, but drops to the 50-52% for *Rhodopseudomonas palustris* strains (genus of α-proteobacteria), 44-47% for *Bradyrhizobium* strains (α-division) and down to 40% for the *Burkholderia xenovorans* (β- division). Importantly, *B. xenovorans* is the only species of the many Burkholderia species whose genome sequences have been determined that contains a PA4872 ortholog. This was confirmed by a BLAST search of the Burkholderiaceae group using the *B. xenovorans* as query that showed only lyase sequences (clearly identified by the active site gating loop sequence) at 30-36% identity.

The second PA4872-like copy is found in *Pseudomonas fluorescens* PfO-1 and in *Pseudomonas putida* W619 but in no other reported species. This paralog is expected to have identical substrate specificity in both organisms because the sequence identity between the two is 84%. Based on the species distribution it is evident that PA4872 performs a specialized metabolic function in a select group of Gram-negative bacteria restricted to all Pseudomonads of defined genome sequence (γ-proteobacteria), a few among the many α-proteobacteria, and a single, environmental, β-proteobacteria. Because oxaloacetate and pyruvate are key intermediary metabolites that link numerous pathways

and because carbonate is used as a precursor for biotin mediated caboxylation reactions, the oxaloacetate decarboxylase could have many potential metabolic uses. Indeed, the genome context analysis described below lead us to the conclusion that the Pseudomonads, which exhibit great metabolic diversity, may utilize the oxaloacetate in a variety of processes that require the supply of pyruvate or bicarbonate.

The genome contexts of the PA4872 and its close relatives were examined to obtain clues about the biological function, underscoring the variety of contexts in which they are found. Only in the *R. palustris strains* and the *B. japonicum* UDSA110 are the oxaloacetate decarboxylase encoding genes organized in an operon. In *R. palustris strains*, the flanking genes (separated by 1-6 nucleotides) are homologs of acyl phosphatase (hydrolysis of acylphosphate metabolites to the corresponding carboxylic acid and orthophosphate) and the α subunit of acyl-CoA carboxylase (phosphorylation of carbonate with ATP to form carboxyphosphate, followed by the transfer of CO_2 from the carboxyphosphate to biotin). We speculate that in these two bacterial species, the bicarbonate produced by the oxaloacetate decarboxylase is fixed in the cell by reaction with the biotin carboxylase subunit of the acyl-CoA carboxylase. The connection with the acyl phosphatase is not clear. However we note that in other bacteria (for example *Nitrobacter hamburgensis* X14) that do not contain the oxaloacetate decarboxylase encoding gene, the genes encoding the acyl phosphatase and the biotin carboxylase are co-transcribed.

The biotin carboxylase, oxaloacetate decarboxylase and acyl phosphatase homologues are also adjacent to one another in the genome of *B. japonicum* UDSA110. However, only the oxaloacetate decarboxylase and acyl phosphatase genes are clearly co-

transcribed (3 nucleotide separation). The biotin carboxylase gene is separated from the oxaloacetate decarboxylase by 149 nucleotides. Perhaps coincidental but nonetheless intriguing, the citrate lyase gene cluster is located in the neighborhood of the biotin carboxylase, oxaloacetate decarboxylase and acyl phosphatase genes within the *R. palustris* BisB18 genome. The citrate lyase converts citrate to oxaloacetate.

The genome contexts of other PA4872 sequence family members are not as informative; the genes are not located on operons and the neighboring genes tend to be different. Yet, it is intriguing that *Pseudomonas stutzeri* A1501, and *Pseudomonas mendocina* ymp, and the two annotated *P. aeruginosa* genomes, PAO1 and PA7, position the oxaloacetate decarboxylase gene in the same neighborhood as the urease operon. Urease, which plays an important role in nitrogen acquisition and recycling requires CO_2 for the synthesis of its nickel metallo center (Park and Hausinger 1995). Next to the urease operon is an amino acid transporter. It is possible that the close arrangement of the amino acid transporter, urease and oxaloacetate decarboxylase genes is not coincidental.

The *P. fluorescens* Pf-5 oxaloacetate decarboxylase gene is included among the genes encoding acetylornithine aminotransferase (of the urea cycle) and the enzymes of the nicotinamide metabolic pathway. *P. fluorescens* PfO-1 has two PA4872 homologs; the closest to PA4872 in sequence (78%) is clustered with arginine N-succinyltransferase, N-formylglutamate amidohydrolase, glutamine synthetase and an amino acid transporter. The respective gene clusters are suggestive of amino acid transport and recycling. The pyruvate generated by the oxaloacetate decarboxylase may serve as a NH_3 acceptor in the synthesis of alanine. The genomes of the *Pseudomonas syringae* strains include the gene encoding oxaloacetate decarboxylase as neighbor to the gene encoding isopropyl malate

synthase (leucine biosynthesis) and the genes encoding an amidase and an aminotransferase.

A common denominator in each of the gene contexts described above is the metabolism of amino acids, in which pyruvate might participate. The oxaloacetate decarboxylase gene of *P. fluorescens* PfO-1 (the second homolog having 55% identity with PA4872) along with that of the *Burholderia xenovans* and *P. putida* F1 are clustered with genes that form a carboxylate transporter and degradation pathway. The genome context of the oxaloacetate decarboxylase gene of *Pseudomonas entomophila* has no relationship to any of the genes listed above, and its neighboring genes share no common function with one another except that there are several genes related to antibiotics resistance.

Based on essential gene analysis in *P. aeruginosa* PAO1 using transposon insertion technology, PA4872 is not required for survival (Jacobs, Alwood et al. 2003). Taken together, the currently available information suggests that the oxaloacetate decarboxylase gene was acquired by all Pseudomonads and by a few other organisms as an auxiliary enzyme that helps maintain the supply of two key metabolites for a variety of cellular processes.

Accession numbers

The coordinates and structure factors have been deposited in the Protein Data Bank (entry code 3B8I).

Chapter 4 Structure and Function of 2,3-Dimethylmalate Lyase, a PEP Mutase/Isocitrate Lyase Superfamily Member

4.1 Abstract

The *Aspergillus niger* genome contains four genes that encode proteins exhibiting greater than 30% amino acid sequence identity to the confirmed oxaloacetate acetyl hydrolase (OAH), an enzyme that belongs to the phosphoenolpyruvate mutase (PEPM)/isocitrate lyase (ICL) superfamily. Previous studies have shown that a mutant *A. niger* strain lacking the OAH gene does not produce oxalate. To identify the function of the protein sharing the highest amino acid sequence identity with the OAH (An07g08390, Swiss-Prot entry Q2L887, 57% identity), we produced the protein in *Escherichia coli* and purified it for structural and functional studies. A focused substrate screen was used to determine the catalytic function of An07g08390 as (2*R*, 3*S*)-dimethylmalate lyase (DMML): k_{cat} = 19.2 s^{-1} and K_m = 220 μM. DMML also possesses significant OAH activity (k_{cat} = 0.5 s^{-1} and K_m = 220 μM). DNA array analysis showed that unlike the *A. niger oah* gene, the DMML encoding gene is subject to catabolite repression. DMML is a key enzyme in bacterial nicotinate catabolism, catalyzing the last of nine enzymatic steps. This pathway does not have a known fungal counterpart. BLAST analysis of the *A. niger* genome for the presence of a similar pathway revealed the presence of homologs to only some of the pathway enzymes. This and the finding that *A. niger* does not thrive on nicotinamide as a sole carbon source suggests that the fungal DMML functions in a presently unknown metabolic pathway. The crystal structure of *A. niger* DMML (in complex with Mg^{2+} and in complex with Mg^{2+} and a substrate analog: the gem-diol of

3,3-difluoro-oxaloacetate) was determined for the purpose of identifying structural determinants of substrate recognition and catalysis. Structure guided site directed mutants were prepared and evaluated to test the contributions made by key active site residues. In this paper we report the results in the broader context of the lyase branch of the PEPM/ICL superfamily to provide insight into the evolution of functional diversity.

4.2 Introduction

The goal of our work is to discover new chemistries catalyzed by the ICL/PEPM superfamily and to determine the unique structural features of the enzyme active sites that diversify the catalytic function of the superfamily active site scaffold. A striking feature of the PEPM/ICL superfamily is the absence of clear sequence boundaries between functionally different enzymes. The high sequence homology between paired enzymes even from different chemical branches (for example 40% identity between carboxyPEPM and the PDP, which is a (2*R*,3S)-2-alkyl-3-methylmalate lyase) has resulted in an unusually high level of miss-annotation of biochemical function in the public sequence data banks. The present study was carried out as part of a larger effort that is aimed at identifying mechanism based sequence markers for use in accurate function assignment. Herein we report on the structure-function study of the unknown protein An07g08390 ORF (GeneBank accession number **DQ349131**, Swiss-Prot entry **Q2L887**) from the oxalic acid producing fungus *Aspergillus niger*. At the outset of this work the *A. niger* genome was known to encode the PEPM/ICL superfamily OAH, which cleaves oxaloacetate to oxalate and acetate. A mutant strain (NW228) lacking the gene does not produce oxalic acid (van den Hombergh, van de Vondervoort et al. 1995; Ruijter, van de Vondervoort et al. 1999). Yet, the *A. niger* genome (Pel, de Winde et al. 2007) contains

four additional OAH homologs, one of which (An07g08390) shares 57% sequence identity with *A. niger* OAH. Although such a high level of sequence identity usually indicates identical function, the phenotype of the gene knockout strain suggests otherwise. Herein, we examine the catalytic function of An07g08390 as defined by *in vitro* substrate screening; the biological function as defined by bioinformatics, gene knockout mutants and DNA array analysis; and the structural basis for the evolution of this novel enzyme within the PEPM/ICL superfamily.

4.3 Materials and methods

4.3.1 Chemicals and General methods

Except where indicated, all chemicals were purchased from Sigma-Aldrich. Oxaloacetic acid, (*R*)-malic acid, (*S*)-malic acid, (2*R*)-methylmalic acid, (2*S*)-methylmalic acid, (2*R*,3*S*)-isocitric acid were purchased from Sigma-Aldrich. (2*R*,3*S*)-2,3-dimethylmalic acid, (2*R*,3*S*)-2-ethyl-3-methylmalic acid, (2*R*,3*S*)-2-propyl-3-methylmalic acid, (2*R*,3*S* & 2*S*,3*R*)-2-methylisocitrate (*threo* isomers) were prepared according to published procedures (Lu, Feng et al. 2005). (3*S*)-Isopropylmalate and 3-butylmalate were a gift from Dr. Steve Clarke, UCLA and CPEP was a gift from Dr. John Gerlt (University of Illinois). Custom-synthesized PCR primers were obtained from Invitrogen as were the restriction enzymes, the *pfu* polymerase, and the T4 DNA lyase. DNA sequencing analysis was carried by the DNA Sequencing Facility of the University of New Mexico. SDS-PAGE was performed with gels prepared from 12% acrylamide gel with a 4% stacking gel (37.5:1 acrylamide:biacryamide ratio) (Bio-Rad). Protein concentrations were determined using the Bradford method.

4.3.2 Cloning, expression and gene product purification

The DMML-encoding An07g08390 gene from the *A. niger* wild type strain N400 (CBS120.49) was amplified by PCR using the forward primer ATGCCTATGGTAACTGCAGCT and the reverse primer CTACTTCAGGTCCACCCCTCC in conjunction with the *Pfu* DNA polymerase kit from Fermentas. The PCR product was cloned in vector pGEM®-T Easy (Promega) and sequenced. The P240T mutant was made by using the Quikchange Site-Directed Mutagenesis method and the mutation was verified by sequencing. The nucleotide sequence of the N400 strain has been deposited in the public databases under accession no **DQ349131**. The DMML-P240T encoding gene was amplified by using a PCR-based strategy (Erlick 1992) with the DMML-P240T-pGEM®-T easy vector clone serving as the template. The Ile142 codon "ATC" was mutated into "ATT" to remove the included *BamH* I restriction site. The clone containing the mutated DMML-P240T gene was digested with *BamH* I and *Ned* I, and the gene ligated to the pET-3c vector (Novagen) cut with *BamH* I and *Ned* I. The recombinant plasmid, DMML-P240T-pET3c, was used to transform competent *E. coli* BL21(DE3) cells (Novagen). DMML wild type was subcloned using Quikchange site-directed mutagenesis method using DMML-P240T-pET3c as the template. The transformed cells BL21 (DE3) were grown at 20 °C with mild agitation (180 rpm) in Luria broth containing carbenicillin (50 µg/mL). After 14 h of cell growth (OD_{600nm} ~ 0.7), induction was initiated with 0.4 mM IPTG (RPI Corp.). The culture was incubated for 10 h at 20 °C under conditions of mild mixing (180 rpm). The cells were harvested by centrifugation (6500 RPM; 7,808*g*) for 15 min at 4 °C in a yield of 3.3 g/ L of culture media. The cell pellet (40 g) was suspended in 400 ml of ice-cold

lysis buffer (50 mM K^+ HEPES (pH 7.5), 1 mM EDTA, 1 mM benzamide hydrochloride, 0.05 mg/ml trypsin inhibitor, 1 mM 1, 10-phenanthroline, 0.1 mM PMSF and 5 mM DTT). The suspension was passed through a French Press at 1200 PSI, and then centrifuged at 4 °C for 1 h at 20000 RPM (48,384*g*). The supernatant was loaded onto a 4.5 x 40 cm Q-Sepharose column equilibrated with 1 L Buffer A (50 mM triethanolamine (pH 7.5), 5 mM $MgCl_2$, 1 mM DTT). The column was washed with 1 L of Buffer A, and then eluted with a 2 L linear gradient of 0 to 0.7 M KCl in Buffer A. The column fractions were analyzed by measuring the absorbance at 280 nm and by carrying out SDS-PAGE analysis. The enzyme eluted at 0.4 M KCl. Ammonium sulfate was added to the combined fractions to generate 25 % saturation. The resulting solution was loaded onto a 3 x 30 cm Butyl Sepharose column equilibrated at 4 °C with 25 % ammonium sulfate in 450 mL of Buffer A. The column was washed with 450 mL of 25 % ammonium sulfate in Buffer A and then eluted with a 1 L linear gradient of 25 % to 0 % ammonium sulfate in Buffer A. The column fractions were analyzed by measuring the absorbance at 280 nm and by carrying out SDS-PAGE analysis. The enzyme eluted at 17 % ammonium sulfate. Ammonium sulfate was added to the combined fractions to generate 25 % saturation. The resulting solution was loaded onto a 3 x 30 cm Phenyl Sepharose column equilibrated at 4 °C with 25 % ammonium sulfate in 450 mL of Buffer A. The column was washed with 450 mL of 25 % ammonium sulfate in Buffer A and then eluted with a 1 L linear gradient of 25 % to 0 % ammonium sulfate in Buffer A. The column fractions were analyzed by measuring the absorbance at 280 nm and by carrying out SDS-PAGE analysis. The enzyme eluted at 0 % ammonium sulfate. The desired fractions were combined, concentrated at 4 °C with an Amicon concentrator and then

dialyzed against Buffer A. The resulting sample was concentrated using a MACROSEP 10K OMEGA for storage at –80 °C. The protein sample was shown to be homogeneous by SDS-PAGE analysis. Yield: 20 mg protein/g wet cell.

4.3.3 Preparation of DMML site-directed mutants

The C124S, C124A, D59S, D59A and Y44F mutant genes were prepared using a PCR-based strategy (Erlick 1992) with *pfu* polymerase, commercial primers and the wild-type-pET3c clone as the template. Mutant protein purifications followed the same protocol as described for the wild-type recombinant enzyme. Protein homogeneity was confirmed by SDS-PAGE analysis.

4.3.4 Preparation of His6-tagged wild-type and P240S mutant enzymes

Owing to the insolubility of the P240S mutant the His_6-tagged protein was prepared. The His_6-tagged wild type enzyme was also prepared to serve as a control. The His_6-tagged proteins were prepared by ligating the *Nde* I and *BamH* I PCR products to the (His_6-tag encoding) pET14b vector linearized with *Nde* I and *BamH* I. The recombinant plasmids were purified and used to transform *E coli.* BL21(DE3) competent cells (Novagen). The transformed cells were grown at 25 °C with mild agitation (180 rpm) in Luria broth (LB) containing 50 μg/mL of carbenicillin. After 11 h of cell growth (OD_{600nm} ~ 0.8), induction was initiated with 0.4 mM IPTG. The culture was incubated for 11 h at 20 °C under conditions of mild mixing (200 rpm). The cells were harvested by centrifugation (6500 RPM (7,808*g*)) for 15 min at 4 °C in a yield of 3.7 g/ L of culture media. The 11 g cell pellet was suspended in 110 mL of ice-cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole and 1 mM DTT, pH 8.0). The suspension

was passed through a French Press at 1200 PSI, and then centrifuged at 4 °C for 60 min at 20000 RPM (48,384*g*). The supernatant was loaded onto a 1.5 × 15 cm Ni-NTA column equilibrated with 200 mL lysis buffer and then rinsed with 200 mL lysis buffer. The column was washed with 300 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 50 mM imidazole and 1 mM DTT, pH 8.0), and then eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole and 1 mM DTT, pH 8.0). The column fractions were analyzed by measuring the absorbance at 280 nm and by carrying out SDS-PAGE analysis. The desired fractions were combined, concentrated at 4 °C with an Amicon concentrator (Amicon) and then dialyzed against Buffer A (50 mM triethanolamine, 5 mM $MgCl_2$, 1 mM DTT, pH 7.5). The resulting sample was concentrated using a MACROSEP 10K OMEGA and stored at –85 °C. The proteins prepared in this manner were shown to be homogeneous by SDS-PAGE analysis. Yield: ~17 mg protein/g wet cell.

4.3.5 Recombinant *A. niger* DMML molecular size determination

The theoretical subunit molecular mass of recombinant DMML was calculated by using the amino acid composition, derived from the gene sequence, and the EXPASY Molecular Biology Server program Compute pI/MW(Appel, Bairoch et al. 1994). The subunit size of recombinant DMML was determined by SDS-PAGE analysis using molecular weight standards from Invitrogen. The subunit mass was determined by electro-spray mass spectrometry (University of New Mexico Mass Spectrometry Lab). The native molecular mass was determined by Sephacryl S-200 gel filtration column (1.5 cm × 180 cm; Pharmacia) chromatography carried out at 4 °C using 0.1 M KCl with Buffer B (50 mM K^+Hepes and 0.5 mM DTT (pH 7.5)) as the eluant. The column was

calibrated using a Pharmacia gel filtration calibration kit (ovalbumin (43 000), aldolase (158 000), ribonuclease A (137 000), and chymotrypsingen A (25 000), albumin (67 000)).

4.3.6 Enzymatic synthesis of (2*R*,3*S*)-dimethylmalic acid by *A. niger* DMML

A 300 mL solution of 40 µM enzyme, 5 mM $MgCl_2$, 5 mM DTT, 500 mM sodium propionate, 500 mM sodium pyruvate, and 100 mM Tris buffer (pH 7.5) was stirred at 25 °C for 3 d. The pH of the solution was adjusted to 10 with 1 M KOH and then hydrogen peroxide (30% aqueous, 30 mL) was slowly added. The resulting solution was stirred at 90 °C for 30 min, and then 10 mg of catalase was added. The mixture was filtered through a 10 kDa membrane (Amicon), concentrated *in vacuo* to ca. 100 mL, adjusted with 12 M HCl to pH 1, and then extracted with ethyl acetate. The ethyl acetate solution was dried over Na_2SO_4 and concentrated *in vacuo* to yield a yellow residue, which was crystallized from ethyl acetate and petroleum ether, yielding 1.2 g of (2*R*,3*S*)-dimethylmalic acid: mp 108-109 °C (reported mp 104-106; (Pirzer, Lill et al. 1979); $[a]^{25}_D$ -13.6 (*c* = 5, H_2O) [reported $[a]^{25}_D$ -16.4 for *c* = 5, H_2O; (Lill, Pirzer et al. 1980); ^{1}H-NMR (500 MHz, *d*-chloroform) d 3.06 (q, *J* = 7.5 Hz, 1H), 1.44 (s, 3H), 1.36 (d, *J* = 7.5 Hz, 3H), 1.25 (s, 1H); 13C NMR d 182.5, 181.2, 74.8, 46.5, 23.6, 10.6; MS (M - 1) *m*/*z* 161.0442, calcd for $C_6H_9O_5$ *m*/*z* 161.0450. Based on the above data, we can conclude that (2*R*,3*S*)-dimethylmalic acid was obtained through the aldol reaction of pyruvate and propionate catalyzed by DMML. The melting point of (2*R*,3*R*)-dimethylmalic acid is 143 °C (Pirzer, Lill et al. 1979).

4.3.7 Activity assays

Continuous lyase assay. The DMML catalyzed C(2)-C(3) bond cleavages in (*R*)-malate, (*S*)-malate, (2*R*)-2-methylmalate, (2*S*)-2-methylmalate, (2*R*,3*S*)-isocitrate, (2*R*,3*S* & 2*S*,3*R*)-2-methylisocitrate (*threo* isomers), (2*R*,3*S*)-isopropylmalate, or 3-butylmalate, (2*R*,3*S*)-dimethylmalate, and (2*R*,3*S*)-2*R*-ethyl-3-methylmalate were continuously monitored spectrophotometrically using a NADH and LDH coupling system. The reaction solution (1 mL) contained 5 mM $MgCl_2$, 20 units/mL lactate dehydrogenase, and 0.2 mM NADH in 50 mM K^+HEPES (pH 7.5 and 25 °C). The absorbance of the solution was monitored at 340 nm (($\Delta\varepsilon$= 6.2 $mM^{-1}cm^{-1}$). The kinetic constants for catalyzed cleavage of (2*R*)-2-methylmalate, (2*R*)-2-ethylmalate, (2*R*,3*S*)-dimethylmalate, and (2*R*,3*S*)-2-ethyl-3-methylmalate and (2*R*,3*S* & 2*S*,3*R*)-2-methylisocitrate were determined using assay solutions in which the reactant concentration was varied from 0.5 K_m to 10 K_m. In the case of (2*R*,3*S*)-2-ethyl-3-methylmalate, reaction solutions contained 600 units of lactate dehydrogenase. To check catalyzed cleavage of (*R*)-malate, (*S*)-malate, (2*S*)-methylmalate, (2*R*,3*S*)-isocitrate, (2*R*,3*S*)-isopropylmalate, or 3-butylmalate malate, kinetic constants were determined with assay solutions in which the reactant concentration is 5 mM and the enzyme concentration is 70 µM. The DMML (150 µM) catalyzed conversion of carboxyPEP (10 mM) to phosphinopyruvate was measured in the presence of 5 mM Mg^{2+} using the assay described in (Lu, Feng et al. 2005).

Fixed-time lyase assay. Reaction solutions initially containing 0.88 µM DMML and 1.5-10 mM (2*R*,3*S*)-2-propyl-3-methylmalate (or 70 µM DMML and 3-15 mM (2*R*,3*S*)-2-isobutyl-3-methylmalate), 5 mM $MgCl_2$, and 50 mM K^+HEPES (pH 7.5, 25 °C) were analyzed at ~20% conversion. A 200-µl aliquot was mixed with 200 µl of 0.2 N

HCl and 100 µl of 0.4 M phenylhydrazine hydrochloride. The solution was stirred for 12 min before the absorbance was measured at either 325 nm (2-propyl-3-methylmalate) or 326 nm (2-isobutyl-3-methylmalate). The standard curve was constructed for each α-ketoacid hydrazone by using the same conditions employed in the enzyme reaction. The molar extinction coefficients for the hydrazones at the wavelength specified above were determined to be 6.8 $mM^{-1}cm^{-1}$ for 2-oxovaleric acid and 6.7 $mM^{-1}cm^{-1}$ for 4-methyl-2-oxopentanoic acid. The control reaction lacked DMML.

OAH assay. OAH activity was assayed according to a published procedure (Lenz, Wunderwald et al. 1976). Reaction solutions (0.5 ml) initially contained 0.14-6 mM oxaloacetate, 7.0 µM DMML, and 5 mM $MgCl_2$ in 50 mM K^+HEPES (pH 7.5 and 25 °C). The reaction was monitored at 255 nm for the disappearance of the enol tautomer of oxaloacetate ($\Delta\varepsilon$= 1.1 $mM^{-1}cm^{-1}$). The rate of oxaloacetate consumption via spontaneous decarboxylation was measured prior to initiating the enzymatic reaction to determine the "background rate". The background rate was subtracted from the reaction rate measured in the presence of DMML. The influence of the buffer properties on the DMML kinetic behavior was tested by replacing the 50 mM K^+HEPES (pH 7.5) of the reaction solutions with 0.1 M imidazole (pH 7.6). The products formed from reaction of 700 µM DMML with 30 mM oxaloacetate in 50 mM potassium phosphate (pH 7.5, 25 °C) containing 5 mM $MgCl_2$ were examined by ^{1}H-NMR. Following incubation for 2 h, the enzyme was removed using a centricon and D_2O was added to a final concentration of 15 % (v/v) before the spectrum was measured with methanol serving as an internal standard (3.39 ppm). The acetate signal was observed as a singlet at 1.96 ppm.

4.3.8 Steady-state kinetic constant determination for *A. niger* DMML substrates

The steady-state kinetic parameters (K_m and k_{cat}) were determined from the initial velocity data measured as a function of substrate concentration. The initial velocity data were fitted to Equation 1 with KinetAsystI,

$$V_0 = V_{max}[S]/(K_m+[S]) \quad (1)$$

where [S] is the substrate concentration, V_0 is the initial velocity, V_{max} is the maximum velocity, and K_m is the Michaelis-Menten constant for the substrate. The k_{cat} value was calculated from V_{max} and the enzyme concentration using the equation $k_{cat}=V_{max}/[E]$, where [E] is the protein subunit molar concentration in the reaction calculated from the ratio of measured protein concentration and the protein molecular mass (31, 986Da).

4.3.9 Determination of inhibition constants for *A. niger* DMML inhibitors

The competitive inhibition constant K_i was determined for 3,3-difluorooxaloacetate (0, 10, 20 μM), oxalate (0, 2, 4, 10 mM) and phosphonopyruvate (0, 5, 10, 20 μM) using (2*R*,3*S*)-2-ethyl-3-methylmalate (1-10 mM) or (2*R*,3*S*)-dimethylmalate (0.1-2.0 mM) as substrate. Assay solutions contained 50 mM K^+HEPES (pH7.5, 25 °C), 5 mM $MgCl_2$,, 600 U/mL LDH and 0.2 mM NADH. The absorbance of the reaction solution was monitored at 340 nm. The competitive inhibition constant K_i was determined by fitting the initial velocity data to Equation 2 with KinetAsystI.

$$V_0 = V_{max}[S]/(K_m(1+(1/K_i))+[S]) \quad (2)$$

4.3.10 Determination of pH-rate profiles for *A. niger* DMML catalysis

The pH dependence of DMML catalysis was measured using (2*R*,3*S*)-2-ethyl-3-methylmalate as the substrate (1-8 mM), 5 mM $MgCl_2$ as the cofactor, and 600 units/mL

LDH and 0.2 mM NADH as the coupling system. The dual buffer system consisting of reaction solutions were buffered at pH 4.5-5.5 with 25 mM acetate and 25 mM MES, pH 5.5-8.0 with 25 mM K^+HEPES and 25 mM MES, pH 8.0-9.0 with 25 mM K^+HEPES and 25 mM TAPS, pH 9.0-9.5 with 25 mM TAPS and 25 mM CAPSO. The initial velocity was fitted with equation 1 to obtain k_{cat} and k_{cat}/K_m, which in turn were used with equation 3 with KinetAsystI to calculate the apparent pK_a values of the ionizing groups.

$$\text{Log } Y=\log(c/1 + [H]/K_1 + K_2/[H]) \quad (3)$$

Where $Y=k_{cat}$ or k_{cat}/K_m, c is the pH-dependent value of Y, [H] is the hydrogen ion concentration, and K_1 and K_2 are dissociation constants of groups that ionize.

4.3.11 Determination of kinetic constants for metal ion activation of *A. niger* DMML

The steady-state kinetic constants for Mg^{2+} and Mn^{2+} activation of DMML for hydrolytic cleavage of oxaloacetate (3 mM) to oxalate and acetate and for cleavage of (2*R*,3*S*)-dimethylmalate (4.7 mM) to pyruvate and propionate were measured with 50 mM K^+-HEPES serving as buffer at pH 7.5 and 25 °C and using assay procedures described above. The initial velocities were measured as a function of Mg^{2+} or Mn^{2+} concentration varied between 0.5-10 fold M^{2+} K_m.

4.3.12 Crystallization and data collection

DMML was crystallized by the vapor diffusion method in hanging drops at room temperature, and in two different crystal forms. The first crystal form contained the enzyme along with Mn^{2+} cofactor and the inhibitor 3,3-difluorooxalacetate. The protein solution contained 10 mg/mL DMML, 10 mM 3,3,-difluorooxalacetate and 5 mM $MnCl_2$.

The reservoir solution consisted of 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane (pH 7.5) and 20% polyethylene glycol (PEG) 3350. The hanging drops contained equal volumes of protein and reservoir solutions. Crystals were obtained within 4 days. The second crystal form was obtained from DMML solution (10 mg/mL) containing 5 mM $MgCl_2$. The reservoir contained 14% PEG 6000 and 0.1 M MES (pH 6.5), and the hanging drops contained equal volumes of protein and reservoir solutions. The crystals appeared within 2-3 days.

For data collection, the DMML/Mn^{2+}/3,3-difluorooxalacetate crystals, immersed in their original mother liquor, were flashed cooled to 100 K in liquid nitrogen. The crystals of the DMML/Mg^{2+} were cryo-protected prior to flash cooling by briefly immersing them in mother liquor containing additional 25% glycerol (in place of water). Diffraction data for both crystal forms were collected at ~100 K on the home facility consisting of Osmic Max-Flux monochromated X-rays generated by a Rigaku Micro Max 007 rotating anode equipped with a Rigaku AXIS IV^{++} image plate detector. Data were processed with CrystalClear/d*trek ((Pflugrath 1999), Rigaku/MSC, the Woodlands, TX). Data collection statistics are shown in Table 4.1.

Table 4.1. Data collection statistics

Sample	DMML-Mn^{2+}-difo	DMML-Mg^{2+}
Wavelength (Å)	1.54	1.54
Space group	P32	P1
Cell dimensions	a = 160.6 Å	a = 79.1 Å
	b= 160.6 Å	b= 115.5 Å
	c = 161.4 Å	c = 115.8 Å
	α = 90˚	α = 119.7˚
	β = 90˚	β = 90.7˚
	γ = 120˚	γ = 96.3˚
Resolution (Å)[a]	2.68 (2.78-2.68)	2.28 (2.36-2.28)
No. observations	303872	321372
No. unique reflections	123774	154242
Completeness (%)[a]	94.7 (90.2)	95.9 (93.8)
R_{merge}[a]	0.111 (0.381)	0.082 (0.324)
Mean I/σI[a]	6.4 (2.0)	5.8 (2.0)
Redundancy[a]	2.5 (2.3)	2.1 (2.1)

[a] $R_{merge} = \sum_{hkl} [(\sum_j | I_j - < I > |) / \sum_j | I_j |]$. Values in parentheses are for the highest resolution shell.

4.3.13 Structure determination and refinement

The structure of the protein/inhibitor complex was determined by Molecular Replacement using the program Phaser (McCoy, Grosse-Kunstleve et al. 2007) as implemented in CCP4 (1994). A partially refined structure of carboxyPEPM determined in our laboratory (unpublished result) was used as the search model. The search model comprised the entire tetramer except for the gating loops, the C-terminal regions spanning the swapped helices and beyond, the bound ligand, and solvent molecules.

Phaser identified the position and orientation of the four tetramers in the asymmetric unit. The gating loop and the C-terminus of one monomer were built manually using the computer graphic program COOT (Emsley and Cowtan 2004), as were a few loop regions that required modification. After completing the building of one molecule, the remaining molecules in the asymmetric unit were generated by applying the non-crystallographic symmetry operators.

Refinement was performed using a maximum likelihood based refinement strategy using the program REFMAC as implemented in CCP4 (1994; Murshudov, Vagin et al. 1997), excluding 5% of the data that were randomly selected for free *R* analysis (Brunger 1992). Refinement of the DMML/inhibitor complex was carried out at 2.6Å resolution. Refinement cycles were alternated with visual inspection of maximum likelihood weighted (2*F*o – *F*c) and (*F*o – *F*c) difference Fourier maps and manual rebuilding with COOT (where *F*o and *F*c are the observed and calculated structure factors, respectively).

The inhibitor, cofactors and the solvent molecules were added during the later stages of refinement. No restraints were imposed on the inhibitor geometry and metal coordination distances. The 3,3-difluorooxalacetate was modeled initially as a ketone, but there was always strong positive density attached to C(2) in the (*F*o – *F*c) difference Fourier map indicating that the bound compound is actually a hydrated C(2) gem diol (the two forms of the compound are depicted in Fig. 4.1B). This positive density persisted throughout the refinement cycles. Towards the end of the refinement, the inhibitor model was changed to gem diol, resulting in an adduct that accounted fully for the electron density (Fig 4.2). At the final stage of the refinement, a TLS cycle was

employed (M.D.Winn 2001) with each of the sixteen molecules treated as an independent unit for fitting T, L and the S displacement tensor parameters.

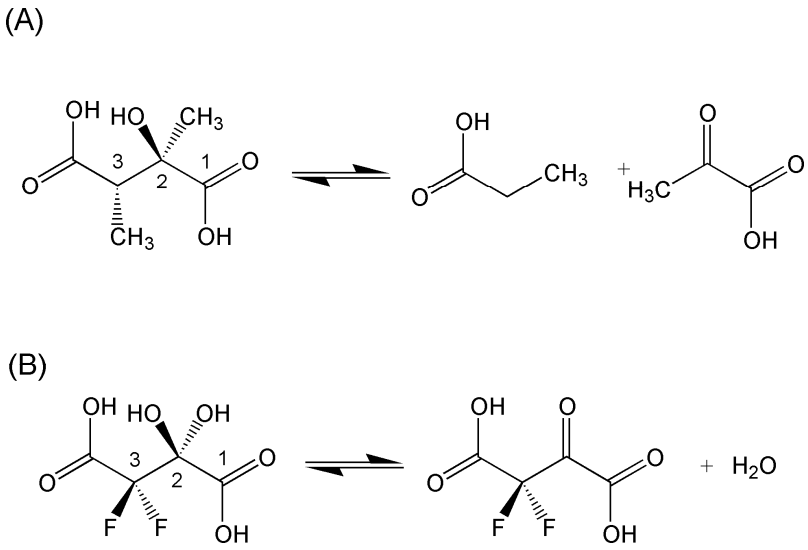


***Figure 4.1**: (A) The DMML-catalyzed reaction, and (B) the two structural forms of 3,3-difuorooxalaceatate. Left – gem diol (major form in solution); Right – ketone (minor form in solution).*

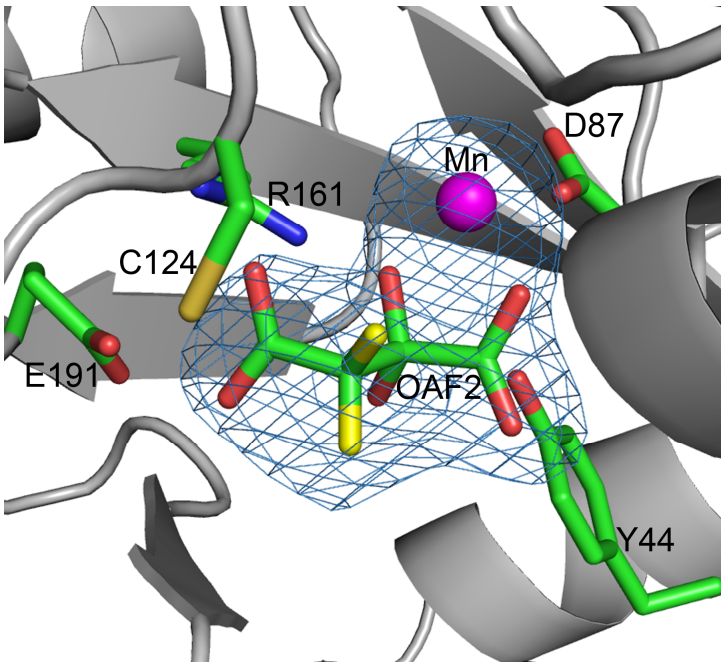


***Figure 4.2**: The $F_o - F_c$ electron density map in the vicinity of the 3,3-difuorooxalaceatate and Mn^{2+} cofactor. The ligand model and the metal cofactor were omitted from the calculation. The map was contoured at 2 σ level. Atomic colors: carbon – green, oxygen – red, nitrogen – blue, fluoride – yellow, Mn^{2+} – magenta*

The crystal structure of the wild type protein with the Mg^{2+} cofactor was also determined by Molecular Replacement using the program PHASER with the DMML/inhibitor structure as the search model (excluding solvent molecules, ligands and the gating loop). This crystal contained three tetramers in the asymmetric unit. Further model building and refinement were carried out following the same protocol as described for the protein/inhibitor complex. The structure statistics are reported in Table 4.2.

Table 4.2. Refinement statistics

Sample	DMML- Mn^{2+}-dfoa	DMML- Mg^{2+}
Resolution (Å)	50 – 2.68	46.8-2.28
No. reflections refined	117564	146395
No. reflections not refined (5%)	6188	7781
$R_{work}/R_{overall}/ R_{free}$ [a]	0.184/0.188/0.256	0.192/0.196/0.261
No. Protein atoms	35456	25205
No. Cofactor (Mn^{2+} /Mg^{2+})	16	12
No. Inhibitor	14	N/A
No. water molecules	704	1162
RMSD from ideal geometry		
Bonds (Å)	0.023	0.022
Angles (°)	2.2	2.1
Mean B value $(Å)^2$	38.5	36.6

[a] $R = \sum_{hkl} | |F_o| - |F_c| | / \sum_{hkl} |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{work} corresponds to the reflections used in refinement, R_{free} refers to randomly selected reflections omitted from the refinement, and $R_{overall}$ corresponds to all reflections.

4.3.14 Structure based sequence alignment

The structure based sequence alignment was generated with VAST (Gibrat, Madej et al. 1996). DMML served as the query structure, incorporating superfamily members with published structure and the unpublished structure of carboxyPEPM determined recently in our laboratory (unpublished result). The OAH sequence was then aligned using CLUSTALW (Pearson and Lipman 1988; Pearson 1990). Finally, a few residues in loop regions required manual adjustment based on inspection of the structure.

4.3.15 Structure analysis

Superposition of structures was carried out within the program COOT (Emsley and Cowtan 2004). The stereo chemical quality of the structures were assessed with the program PROCHECK (Laskowski 1993). The program PyMOL (DeLano 2002) was used for structure depictions.

4.4 Results and Discussion

4.4.1 *A. niger* An07g08390 purification and mass determination

The recombinant *A. niger* An07g08390 was purified to homogeneity in a yield of 20 mg/g wet cells by using a column chromatography-based protocol. The theoretical mass of An07g08390 is 32,117 Da whereas that determined by mass spectrometry is 31,986 Da. Thus, the N-terminal Met is removed by posttranslational modification. The SDS-PAGE analysis gave an estimated subunit mass of 32 kDa, whereas the native mass measured using molecular size gel filtration chromatography is ~110 kDa, consistent with the tetrameric structure observed in the crystal structure.

4.4.2 Substrate screen of *A. niger* An07g08390

The first objective was to identify potential candidates for the physiological substrate of the *A. niger* enzyme An07g08390 by using a focused substrate-activity screen. The high sequence identity with the *A. niger* OAH, and the presence of two key lyase catalytic residues (*viz.* the active site gating loop cysteine presumed to function in general acid catalysis, and the protonated active site glutamic acid residue presumed to function in electrophilic catalysis (Sharma, Sharma et al. 2000; Liu, Lu et al. 2005)), which specifically serve the lyase branch of the superfamily directed our attention to α-hydroxyacid metabolites as potential substrates. We began the substrate screen with the substrates of other known lyase branch members, namely ICL, MICL, PDP (the preferred PDP substrate is (2*R*,3*S*)-2-alkyl-3-methylmalate), and OAH. Isocitrate is not a substrate and 2-methylisocitrate is a very poor substrate (Table 4.3). The oxaloacetate is cleaved to oxalate and acetate presumably via the enzyme bound C(2)gem diol. Because the ratio of ketone to gem diol under the reaction conditions is 13:1 (Lu, Feng et al. 2005), the actual k_{cat}/K_m value might be as large as 3×10^4 M^{-1} s^{-1}. The (2*R*,3*S*)-2-alkyl-3-methylmalates (2*R*,3*S*)-dimethylmalate ($k_{cat}/K_m = 9 \times 10^4$ M^{-1} s^{-1}) and (2*R*,3*S*)-2-ethyl-3-methylmalate ($k_{cat}/K_m = 5 \times 10^4$ M^{-1} s^{-1}) are slightly better substrates. (*R*)-malate, (*S*)-malate, (2*S*)-methylmalate, (3*S*)-isopropylmalate and 3-butylmalate are not substrates. Likewise, AN07G08390 showed no detectable activity associated with the other family members including PEPM, phosphonopyruvate hydrolase, and carboxyPEPM. Based on this substrate screen, and the structural and *in vivo* data reported in later sections, we name the *A. niger* enzyme (2*R*,3*S*)-dimethylmalate lyase or simply "DMML". DMML's assigned biochemical reaction is shown in Fig. 4.1A.

Table 4.3. Steady-State Kinetic and Competitive Inhibition Constants Determined for *A. niger* DMML in 5 mM $MgCl_2$ and 0.1 M imidazole (pH 7.6 and 25 °C) or 5 mM $MgCl_2$ and 50 mM K^+HEPES (pH 7.5 and 25 °C).

Reactant	k_{cat}(s^{-1})	K_m(mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	K_i[f]
Oxaloacetate[a]	0.505 ± 0.005	0.234 ± 0.008	2 x 10^3	------
oxaloacetate [b]	0.48 ± 0.02	0.22 ± 0.03	2 x 10^3	------
2R-methylmalate[c]	2.64 ± 0.09	2.2 ± 0.2	1 x 10^3	------
2R-ethylmalate [d]	4.4 ± 0.7	12 ± 3	1 x 10^2	------
2R, 3S-dimethylmalate[c]	19.2 ± 0.5	0.22 ± 0.02	9 x 10^4	-------
2R-ethyl-3S-methylmalate[d]	63 ± 2	1.17 ± 0.09	5 x 10^4	-------
2R-propyl-3S-methylmalate[e]	0.79 ± 0.01	2.7 ± 0.1	3 x 10^2	-------
2R, 3S & 2S, 3R-2-methylisocitrate[c]	0.13 ± 0.01	12 ± 3	1 x 10^1	-------
2R, 3S-isocitrate[c]	<1 x 10^{-5}	------	------	>5 mM[h]
R-malate[c]	<1 x 10^{-5}	------	------	>5 mM[h]
S-malate [c]	<1 x10^{-5}	------	------	>5 mM[h]
2S-methylmalate[c]	<1 x 10^{-5}	------	------	-------
3S-isopropylmalate[c]	<1 x 10^{-5}	------	------	-------
3-butylmalate[c]	<1 x 10^{-5}	------	------	-------
phosphonopyruvate [f]	<1 x 10^{-5}	------	------	2.4 ± 0.1 µM[h]
carboxyPEP[g]	<1 x 10^{-5}	------	------	>5 mM[h]
3,3-difluoroxaloacetate	<1 x 10^{-5}	------	------	2.5 ± 0.2 µM[h]
oxalate	-----	------	------	1.6 ± 0.1 mM[h]

[a] The kinetic constants were determined using direct optical method and imidazole as buffer. [b] The kinetic constants were determined using direct optical method and K^+HEPES as buffer. [c] The kinetic constants were determined using LDH/NADH coupling assay and (20 units/mL LDH) K^+HEPES as buffer. [d] The kinetic constants were determined by LDH/NADH coupling assay (600 units/mL LDH) K^+HEPES as buffer.[e] The kinetic constants were determined using the fixed-time phenylhydrazine-based assay K^+HEPES as buffer. [f]The mutase activity was monitored using the pyruvate kinase/LDH/NADH coupling assay and K^+HEPES as buffer and the hyrolase activity was monitored using the LDH/NADH coupling assay and K^+HEPES as buffer. [g]The malate dehydrogenase/NADH coupling assay and K^+HEPES buffer was used. [h]The K_i value was determined using (2*R*,3*S*)-dimethylmalate or (2*R*,3*S*)-2-ethyl-3methylmalate as substrate and K^+HEPES as buffer .

4.4.3 DMML inhibitors

In order to determine binding specificities (addressed in a later section), and identify a suitable active site ligand for DMML structure determination, a select group of compounds were screened as competitive inhibitors (Table 4.3). (2*R*,3*S*)-isocitrate, (*R*)-malate, (*S*)-malate and carboxy-PEP are not inhibitors (K_i > 5 mM). Phosphonopyruvate and 3,3-difluorooxaloacetate are tight binding competitive inhibitors (K_i = 2.4 and 2.5 μM, respectively) whereas oxalate is a weak binding competitive inhibitor (K_i = 1.6 mM). The 3,3-difluorooxaloacetate, whose structure is shown in Fig. 4.1B, was used in co-crystallization with DMML.

4.4.4 Buffer, metal ion and pH effects on DMML catalysis

In order to determine the optimal reaction conditions for DMML catalysis studies were carried out to test buffer, metal ion and pH effects. The pH rate profiles measured for DMML are shown in Fig. 4.3. The k_{cat} and k_{cat}/K_m values were determined at varying concentrations of (2*R*,3*S*)-2-ethyl-3-methylmalate and 5 mM $MgCl_2$ as a function of the reaction solution pH. The plot of log (k_{cat}) *vs* pH showed that the maximum activity of DMML extends from pH 4.5-7.5 and drops at higher pH (apparent pK_a = 8.7 ± 0.1). The plot of log (k_{cat}/K_m) *vs.* pH is bell-shaped, indicating a loss of activity at both acidic (apparent pK_a = 6.1 ± 0.3) and basic pH (apparent pK_a value of 8.3 ± 0.3). Therefore, all kinetic determinations were carried out at pH 7.5 where DMML displays optimal activity. DMML catalysis involves proton transfers between several residues. The protonation states of these residues are thus interconnected. Consequently, no attempt is made to assign the apparent pK_a values to specific enzyme or substrate groups.

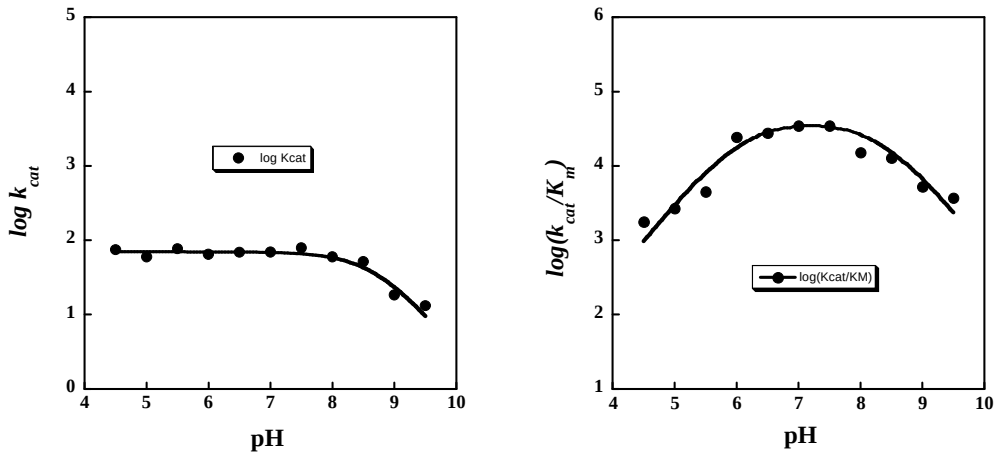


Figure 4.3: *The pH-rate profiles for the DMML-catalyzed reaction of (2R)-ethyl-(3S)-methylmalate measured in the presence of 5 mM $MgCl_2$ at 25 ℃. See Materials and Methods for details.*

The pH rate profiles described above were measured using MES, HEPES, TAPS, CAPSO and acetate. No buffer effects were observed. Likewise, the kinetic constants for DMML catalyzed cleavage of (2*R*,3*S*)-dimethylmalate measured using imidazole or HEPES (Table 4.3) are equivalent. The k_{cat} and K_m for Mg^{2+} activation of DMML catalyzed cleavage of (2*R*,3*S*)-dimethylmalate (4.7 mM) in 50 mM K^+HEPES (pH 7.5, 25 ^oC) are 10.7 ± 0.4 s^{-1} and 22 ± 2 μM, respectively. The k_{cat} and K_m for Mg^{2+}activation of DMML catalyzed cleavage of oxaloacetate (3 mM) in 50 mM K^+HEPES (pH 7.5, 25 ^oC) are 0.39 ± 0.01 s^{-1} and 21 ± 2 μM, respectively. The k_{cat} and K_m for Mn^{2+} activation of DMML catalyzed cleavage of oxaloacetate (3 mM) in 50 mM K^+HEPES (pH 7.5, 25 ^oC) are 0.92 ± 0.01 s^{-1} and 8.1 ± 0.7 μM, respectively. Mn^{2+} and Mg^{2+} were used in the DMML crystal structure determinations.

4.4.5 Overall structure of *A. niger* DMML

Quaternary Structure: The DMML/Mg^{2+} crystal structure was refined at 2.3 Å resolution (Table 4.2). The asymmetric unit contains 12 subunits assembled into 3 tetramers. The DMML/ Mn^{2+}/3,3-difluorooxalacetate crystal structure was refined at 2.7 Å resolution (Table 4.2; Fig. 4.4). The asymmetric unit contains sixteen subunits assembled into four tetramers. As with the structures of most of the known PEPM/ICL superfamily members, the tetramers are comprised of dimer of dimers with each subunit adopting the $(\alpha/\beta)_8$ barrel fold. Two subunits associate to form the dimer by swapping the eighth α-helix of the barrel and the two dimers associate to form the biological tetramer (Fig. 4.4).

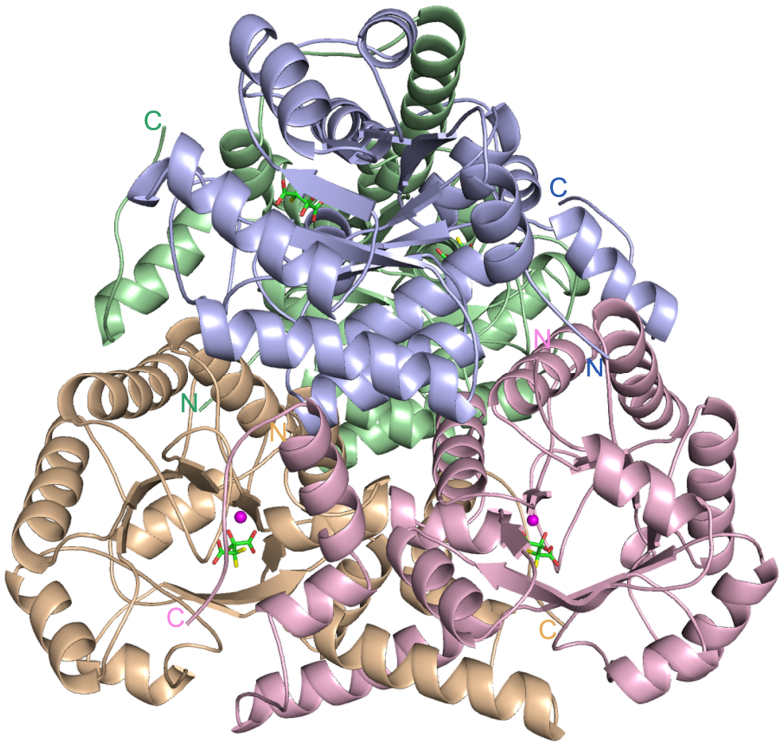


Figure 4.4: *The DMML tetramer structure, highlighting each subunit in different color. The gold and salmon molecules swap their $(\alpha/\beta)_8$ barrel 8th α-helices, as does the pair of blue and green molecules. Mn^{2+} is depicted as magenta sphere and the inhibitor as stick models with the atomic coloring scheme: carbon - green, oxygen – red, fluoride-yellow.*

Active Site Gating Loop: The active sites of most known PEPM/ICL superfamily members are desolvated by the C-terminus of the opposing subunit and by a flexible active site gating loop. All twelve subunits of the DMML/Mg^{2+} crystal structure contain a Mg^{2+} bound to the active site. The active site gating loop, formed by residues 121 - 130, is disordered in all but one subunit (molecule L in the coordinates deposited in the PDB). In this subunit the loop adopts a well-defined "open" conformation stabilized by inter-tetramer contacts (the open conformation is shown in Fig. 4.5A together with the loop-closed conformation of a DMML molecule bound with 3,3-difluorooxalacetate and Mn^{2+}). The DMML/Mn^{2+}/3,3-difluorooxalacetate crystal structure reveals an ordered "closed" active site gating loop conformation (Fig. 4.5A) for the fourteen subunits, which contain active site bound Mn^{2+} and 3,3-difluorooxalacetate. The loop residues 121-130 of the two subunits (J and N in the coordinates deposited in the PDB) that contain active site bound Mn^{2+} but no 3,3-difluorooxalacetate ligand, are disordered.

The C-terminal Segment; Six to ten C-terminal residues of the subunits of the DMML/Mg^{2+} crystal structure are disordered. On the other hand, the C-terminal segments of the subunits observed in the DMML/Mn^{2+}/3,3-difluorooxalacetate crystal structure are ordered. Two different C-terminal segment conformations are evident in the enzyme/inhibitor complex. The conformation observed in all subunits except for I and M allows the C-terminal segment to cover a portion of the entrance to the active site of the paired subunit. The remaining active site entrance is covered by the gating loop that has assumed the closed conformation. The conformation of the C-terminal segment of subunits I and M directs the peptide away from the active site of the paired subunit, thus exposing it to solvent (Fig. 4.5B).

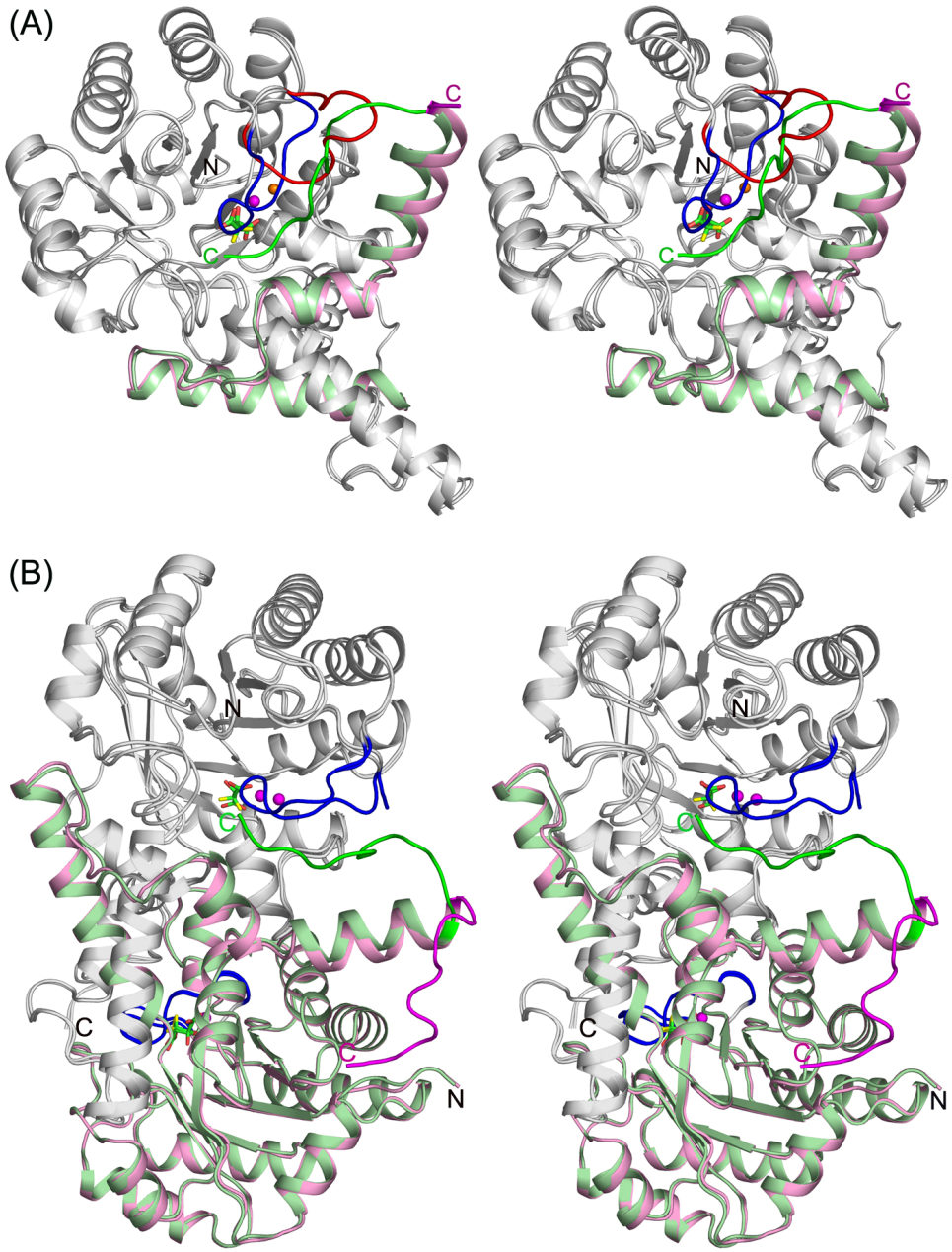


Figure 4.5*: Stereoscopic views of superposed DMML molecules depicting conformational flexibility. (A) The open gating loop (molecule L of the DMML/Mg^{2+} structure) and the C-terminus of the loop-closed state (molecule A of the DMML/Mn^{2+}/3,3-difuorooxalaceatate structure) occupy the same space, implying that the transition requires correlated movement. The open gating loop is colored red, the closed gating loop is colored blue. The C-terminal region (residues 292-301) of molecule A is highlighted in green. The same region is disordered in molecule L, where the last two ordered residues are highlighted in magenta. Mn^{2+} is shown as a orange sphere and Mg^{2+} is shown as a magenta sphere. The metal co-factor shifts 2.3 Á upon inhibitor binding. The 3,3-difuorooxalaceatate is shown in atomic colors (carbon-green, oxygen-red, fluoride-yellow). (B) Two different C-terminus conformations in the crystal structure of DMML/Mn^{2+}/3,3-difuorooxalaceatate complex. Molecule M (whose dimeric partner lacks inhibitor in the active-site) is superposed on molecule A (from a dimer occupied with inhibitor in both subunits). Mn^{2+} and atomic color are as in 6A. Gating loops are shown in blue. The C-terminal region (residues 292-*

301) is highlighted in red and green in molecules M and A, respectively. The polypeptide traces of the two corresponding molecules are highlighted pale red and pale green.

Notably, the respective subunits that pair with I and M are J and N respectively. Subunits J and N differ from the other subunits in that their flexible gating loops (residues 121-130) assume the "open" rather than the "closed" conformation. The "open" loop conformation prevents the C-terminal segment of the paired subunit from binding at the active site entrance. Thus, the active sites of subunits J and N are fully solvent exposed whereas the active sites of all of the other subunits are fully closed to solvent.

To illustrate this unique feature, a subunit having an ordered, open gating loop conformation defined in the crystal structure of the DMML/Mg^{2+} was superimposed with a subunit bound with Mn^{2+} and 3,3-difluorooxalacetate and having an active site gating loop in the closed conformation (Fig. 4.5A). The superposition reveals that part of the open gating loop occupies the same space occupied by the C-terminus of a molecule in the closed conformation. The correlated movement of the gating loop and the C-terminus is thus necessary to avoid clashes. Correlated movement of the C-termini and the gating loop is also seen in ICL (Sharma, Sharma et al. 2000). MICL exhibits a different mode of correlated movement (Liu, Lu et al. 2005).

4.4.6 DMML active site structure

As with other α/β barrel enzymes (Branden 1980), the DMML active site is located at the C-terminal end of the barrel's β-strands. In this section we use the DMML/Mn^{2+}/3,3-difluorooxalacetate crystal structure to identify the active site residues that interact with the divalent metal ion cofactor and the substrate ligand (Fig. 4.6A). The 3,3-difluorooxalacetate ligand is bound to the active site as the C(2) gem-diol. The gem-

diol, which is formed by hydration of the 3,3-difluorooxalacetate C(2)=O (Fig. 4.1B), is the major structural form observed in aqueous solution (Han, Joosten et al. 2007). Thus, one C(2)OH ("reacting C(2)OH") assumes the position of the C(2)OH in the (2*R*,3*S*)-dimethylmalate (Fig. 4.1A) and the position of the Pro-*R* C(2)OH in the oxaloacetate C(2)gem-diol.

During catalytic turnover of (2*R*,3*S*)-dimethylmalate, (2*R*,3*S*)-2-ethyl-3-methylmalate and oxaloacetate C(2) gem-diol, the reacting C(2) hydroxyl group undergoes deprotonatation in concert with, or prior to, C(3)-C(2) bond cleavage. The other 3,3-difluorooxaloacetate C(2)OH ("binding C(2)OH") assumes the position of the $C(2)CH_3$ in (2*R*,3*S*)-dimethylmalate and the Pro-*S* C(2)OH in the oxaloacetate C(2)gem-diol. The reacting C(2)OH coordinates to the Mn^{2+} and engages in hydrogen bond formation with the Arg161 guanidinium group, whereas the binding C(2)OH engages in hydrogen bond formation with the Tyr44 hydroxyl group. The C(1)OO- uses one oxygen atom to coordinate to the Mn^{2+} and to form a bifurcated hydrogen bond with the backbone amide NHs of Gly47 and Ala48. The other C(1)OO- oxygen atom engages in hydrogen bond interaction with the Thr46 hydroxyl group. The C(4)OO- forms hydrogen bonds with the Asn214 side chain, the protonated Glu191 side chain, the backbone amide NH of Gly125 and the side chain of Arg161. In addition to the two coordination bonds formed with the 3,3-difluorooxaloacetate C(1)OO- and reacting C(2)OH, the Mn^{2+} coordinates the Asp87 carboxylate group and three water molecules to form an octahedral geometry complex (average metal-ligand distance of 2.1 Å). The three water ligands engage in hydrogen bond formation with the COO groups of Asp59, Asp89 and Glu116.

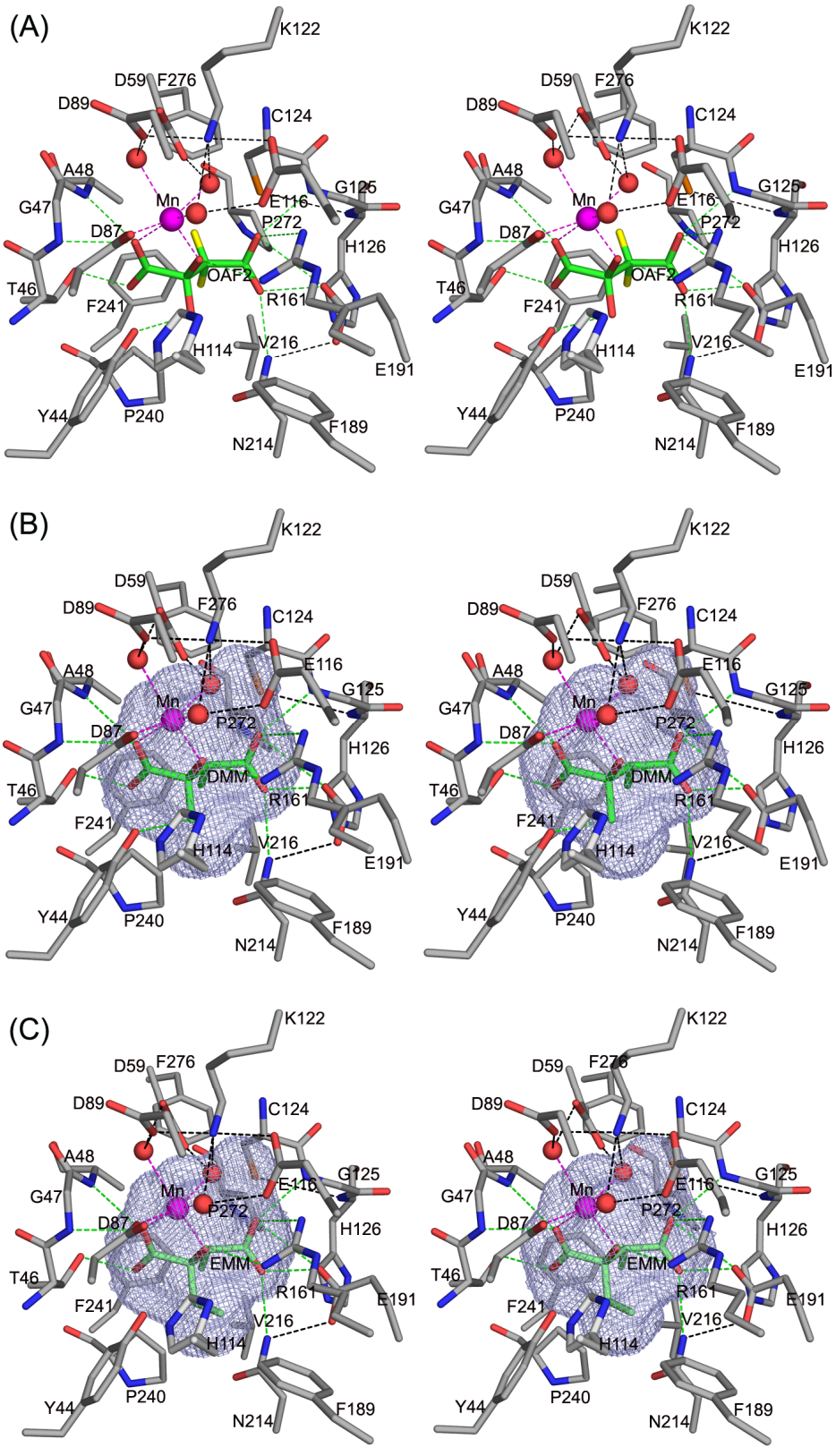

(A)
K122
D59
F276
D89
C124
A48
G125
G47
Mn
E116
P272
D87
H126
OAF2
T46
F241
R161
H114
V216
E191
Y44
P240
F189
N214
(B)
DMM
(C)
EMM

***Figure 4.6**: DMML active site. (A). Stereoscopic representation of Mn^{2+}/3,3-difluorooxalacetate binding site. Atomic colors are as in Figure 4.4 except that protein carbon atoms are colored grey. Key interactions are highlighted with dotted lines as follows: Coordination to Mn^{2+} is shown in magenta, electrostatic interactions with the ligand are shown in green, and protein interactions with Mn^{2+}-coordinated water molecules and protein-protein interactions are shown in black. (B.) Stereoscopic representation of the active site modeled with 2R, 3S-dimethylmalate in place of the 3,3-difluorooxalacetate ligand. The cage represents the space vacated by removing the 3,3-difluorooxalacetate and the metal cofactor. (C.) Stereoscopic representation of the active modeled with 2R-ethyl,3S-methylmalate in place of the 3,3-difluorooxalacetate ligand.*

In the absence of the bound 3,3-difluorooxaloacetate only the direct coordination of Asp87 carboxylate group to the Mn^{2+} is evident. The metal coordination in the DMML/Mg^{2+} structure is somewhat different. Here both Asp87 and Asp89 are directly coordinated to Mg^{2+} in all but one molecule in the asymmetric unit (molecule L in the structure deposited in the PDB), wherein only Asp89 serves as direct ligand to Mg^{2+} and a water molecule bridges the Asp87 carboxylate group. The average Mg^{2+}-ligand distance is 2.3 Å. A complete octahedral coordination is seen only in one molecule (D). The displacement of the metal cofactor position upon ligand binding is on average 2.3 Å, comparable with the shifts reported previously for MICL (2.7 Å) and ICL (2.4 Å) (Sharma, Sharma et al. 2000; Grimm, Evers et al. 2003; Simanshu, Satheshkumar et al. 2003; Liu, Lu et al. 2005).

In order to gain insight into the structure of the enzyme-substrate complex, (2*R*,3*S*)-dimethylmalate was modeled into the DMML active site in place of the 3,3-difluorooxalacetate ligand. This model (Fig. 4.6B) conserves the interactions observed in the DMML/Mn^{2+}/3,3-difluorooxalacetate crystal structure and defines the positions of the *R*-C(2) and *S*-C(3) methyl groups relative to the surrounding active site residues. Interestingly, the region surrounding the *S*-C(3)H would also appear to accommodate a methyl group. This raised the possibility that the (2*R*,3*R*)-dimethylmalate might also be a substrate. The (2*R*,3*S*)-dimethylmalate was synthesized by using the PDP to catalyze the

reaction of pyruvate and propionate and then purifying the dimethylmalate product and showing that it has the (2*R*,3*S*) configuration (Lu, Feng et al. 2005). The same procedure was carried out with DMML substituting for the petal death protein with the same result (see Materials and Methods for details). Thus, the preferred substrate for both enzymes is the known metabolite (2*R*,3*S*)-dimethylmalate. (2*R*,3*R*)-Dimethylmalate, on the other hand, is not a known metabolite. We will return to the (2*R*,3*S*)-dimethylmalate model in the discussion of the mechanism of substrate recognition and catalysis in the following section.

Residues that interact with the substrate's α-ketoacid or α-hydroxyacid moiety are conserved in the PEPM/ICL superfamily (Fig. 4.7). The active site residue, Lys122, involved in loop opening and closure (Liu, Lu et al. 2004) is also conserved in other PEPM/ICL superfamily members except for OAD (Narayanan, Niu et al. 2008). Lys122 interacts with Asp89 and Glu116, and is protected from unfavorable contact with the Mn^{2+} by two of the three Mn^{2+} water ligands.

Residues defining lyase activity within the PEPM/ICL superfamily are conserved in DMML (Fig. 4.6A and 4.7): the catalytic cysteine (Cys124 in DMML), the likely proton donor to C(3); Glu191 that shares a proton with the C(4)OO-; and Asn214 that anchors Glu191 appropriately for its role in electrophilic catalysis. As observed in MICL, ICL, and PDP, the Cys124 thiol group interacts with the backbone amide of His126, thus forming an Asx turn. This interaction is expected to decrease the pK_a of the cysteine and aid in proton transfer to the ("aci-carboxylate") carbanion intermediate that develops upon C(3)-C(2) bond cleavage (Liu, Lu et al. 2005).

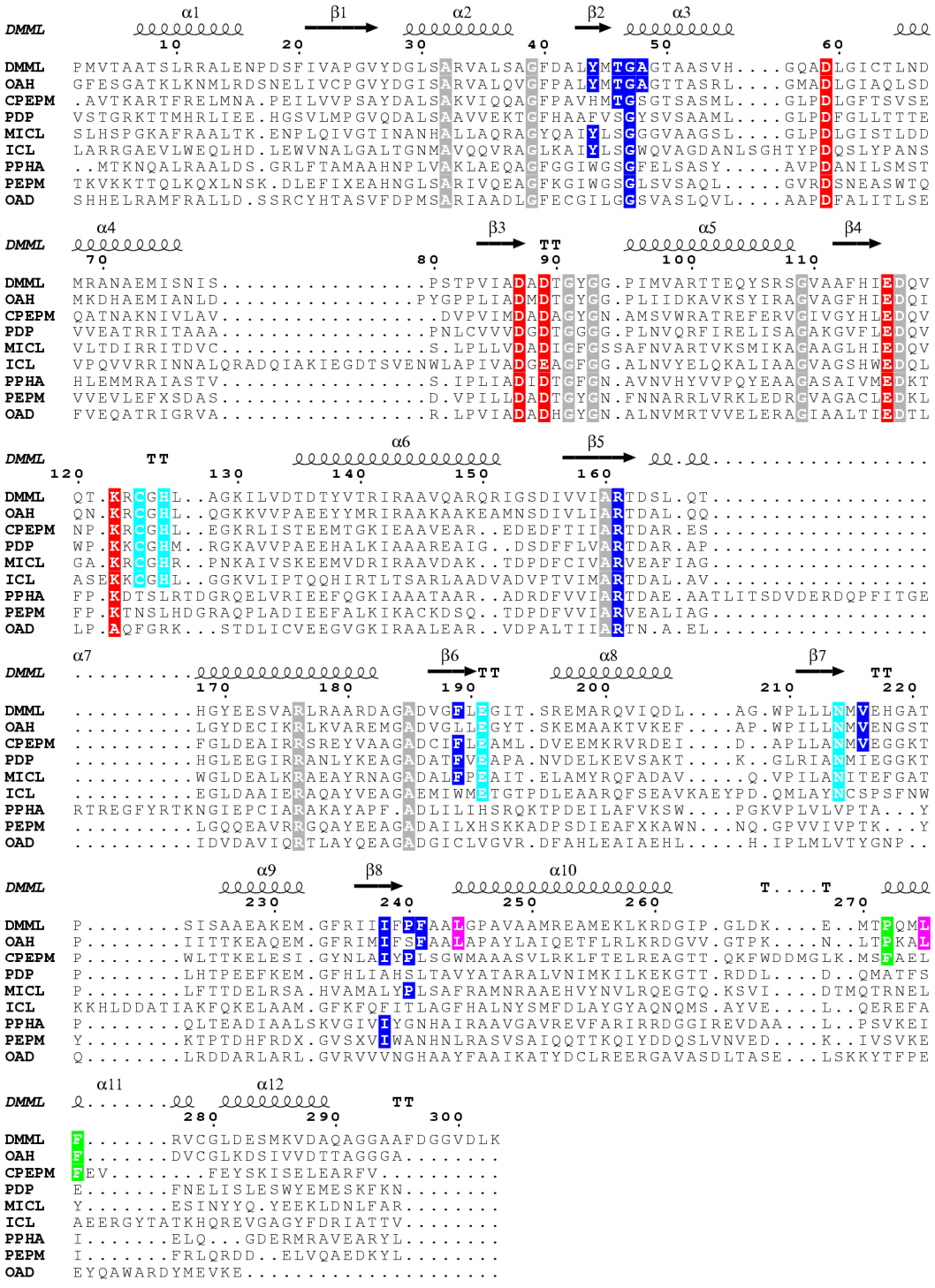


***Figure 4.7**: Structure based sequence alignment of PEPM/ICL superfamily members. Mg^{2+}/Mn^{2+} coordinating residues are shown in red. DMML residues that are coordinated to the inhibitor and/or line the active site are shown in blue as are conserved counterparts of other homologs. The lyase signature residues are shown in cyan. The two leucine residues that undergo correlated conformational change together with Phe241 are shown in pink. C-terminal residues that line the active site of a partner molecule are shown in green. Other invariant residues are shown in gray.*

In the previous section we noted correlated movements between the gating loop and the C-terminus of DMML. Correlated shifts are also observed locally in response to the 3,3-difluorooxaloacetate ligand, affecting close-by residues including a residue of the partner subunit. The Phe241 side chain adjusts to enhance van der Waals contact with the C(3)F, which suggests that the Phe241 ring interacts with the C(3) methyl group of the substrate. The ligand induced change in the Phe241 conformation is accompanied by changes in the conformations of neighboring residues, Leu244 of the same subunit and Leu275 of the adjacent subunit, suggesting coupled motion (Fig. 4.8).

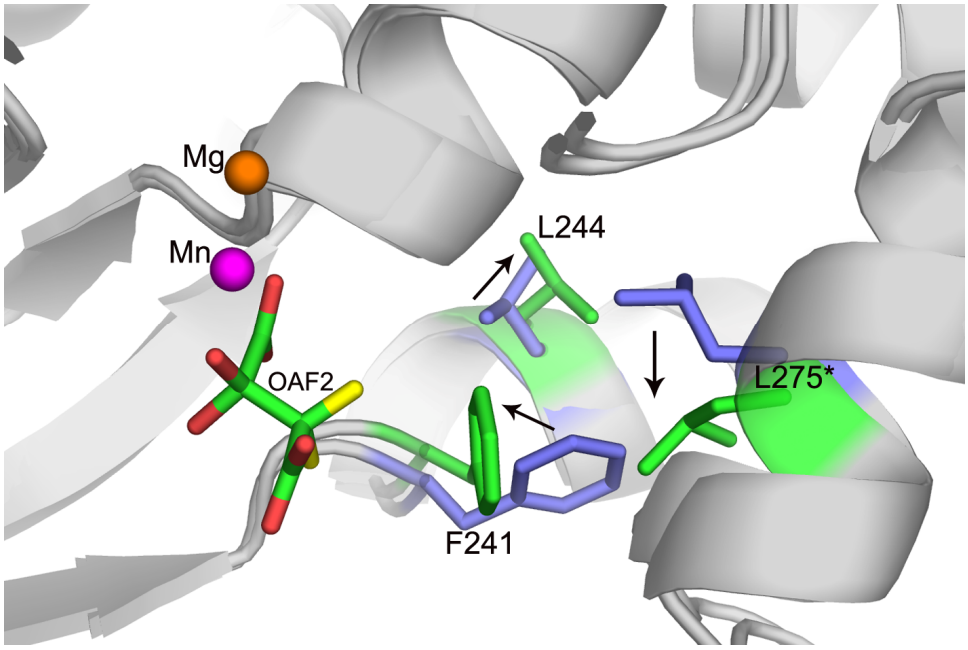


***Figure 4.8**: Coordinated movement upon ligand binding. The residues of the DMML/Mn^{2+}/3,3-difuorooxalaceatate complex are shown in green. The residues of the DMML/Mg^{2+} complex are shown in blue. Mn^{2+} and Mg^{2+} are shown as magenta and orange spheres, respectively. Atomic coloring of the inhibitor is as in figure 4.4. The direction of movement of residues upon ligand binding is indicted by black arrows.*

4.4.7 Structural basis for substrate specificity in DMML

The substrate and inhibitor screens show that DMML targets the R-malate frame, yet it does not catalyze the cleavage of R-malate (Table 4.3) nor, remarkably, does it bind R-malate (Ki >20 mM). What are the structural determinants that require substitutions at C(2) and C(3)?

Recognition of the C(2) substituent: Based on earlier structure determinations of ICL, MICL, and PDP (a (2*R*,3*S*)-2-alkyl-3-methylmalate lyase which also possesses comparable OAH activity) it has been suggested that a triad of active residues forms a pocket that defines the binding site for 2*R*-C(2) substituent. In MICL, the triad is Phe186, Leu234, and Pro236; in ICL the analogous residues are Trp283, Phe345, and Thr347; and in the PDP the residues are Phe207, Ala255 and Ser257 (Grimm, Evers et al. 2003; Liu, Lu et al. 2005; Teplyakov, Liu et al. 2005). MICL can accommodate a methyl group in the pocket, ICL can only accommodate a hydrogen atom, and PDP accommodates hydroxyl, methyl, ethyl, or even a propyl group at the 2*R*-C(2) position (Lu, Feng et al. 2005). The polar interaction between PDP Ser257 and the pro-*S* hydroxyl group of the oxaloacetate gem-diol is suggested by structural (Teplyakov, Liu et al. 2005) and mutagenesis (Joosten, Han et al. 2008) data to account for PDP's ability to function equally well as a (2*R*,3*S*)-2-alkyl-3-methylmalate lyase and an OAH. For DMML, the "specificity triad" consists of Phe189, Ile238, and Pro240, and thus matches that of the MICL (Fig. 4.6 & 4.7), and similarly, delineating a pocket for a 2*R*-C(2) methyl group. A model of the DMML active site bound with (2*R*,3*S*)-2-ethyl-3-methylmalate (Fig. 4.6C) indicates that the C(2) ethyl group can be accommodated as well, consistent with the observation that (2*R*,3*S*)-2-ethyl-3-methylmalate is a good substrate of DMML (k_{cat}/K_m =

5 x 10^4 $M^{-1}s^{-1}$ *vs*. 9 x 10^4 $M^{-1}s^{-1}$ for (2*R*,3*S*)-dimethylmalate; Table 4.3). The larger 2*R*-propyl group on the other hand is not accommodated well as evidenced by the 300-fold drop in the k_{cat}/K_m value measured for (2*R*,3*S*)-2-propyl-3-methylmalate (Table 4.3). The ability of DMML to catalyze hydrolytic oxaloacetate cleavage (Table 4.3) can perhaps be explained by the polar interaction between the Tyr44 and the pro-*S* hydroxyl group of the oxaloacetate gem-diol as is suggested by the structure of the DMML/Mn^{2+}/3,3-difluorooxaloacetate (Fig. 4.6A). Tyr44 does not correspond spatially to the Ser257 of the PDP, but rather its sequence counter part in PDP is Phe64. Replacement of the DMML Tyr44 with Phe by site-directed mutagenesis removed all detectable activity towards (2*R*)-2-methylmalate, (2*R*,3*S*)-dimethylmalate, (2*R*,3*S*)-2-ethyl-3-methylmalate and oxaloacetate (Table 4.4). For MICL, we proposed that the Tyr43 (counterpart of OAH Tyr44) hydroxyl group (which is hydrogen bonded to His113, also conserved in OAH) may form part of the catalytic machinery (Liu, Lu et al. 2005). The elimination of all catalytic activities in the DMML Y44F mutant supports this proposal as does the conservation of the Tyr in most of the lyases within the PEPM/ICL superfamily. The catalytic activity of PDP, which contains a Phe-Phe pair instead of the Tyr-His pair might be rationalized by the presence of an internal water molecule that plays a similar role to the hydroxyl group of the Tyr residue in all other lyases. The PDP Phe and its His counterpart in DMML and MICL are oriented differently such that the PDP contains a niche for a water molecule in the same position that is occupied by the His in DMML and MICL.

Table 4.4. Steady-State kinetic constants determined for the *A. niger* DMML site-directed mutant mutants for the catalyzed conversion of (2*R*,3*S*)-dimethylmalate (DMM) to pyruvate and propionate, or the catalyzed conversion of oxaloacetate (OA) to acetate and oxalate, 50 mM K^+HEPES buffer (pH 7.5 and 25˚ C) containing 5 mM $MgCl_2$.

DMML	substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
WT	DMM	19.2 ± 0.5	220 ± 20	9×10^4
Y44F[a]	DMM	< 1 x 10^{-5}	-------	----
D59A	DMM	(4.27 ± 0.09) ×10^{-4}	7400 ± 400	6×10^{-2}
D59S	DMM	(1.4 ± 0.1) ×10^{-4}	6000 ± 1000	2×10^{-2}
C124A	DMM	(1.78 ± 0.05) ×10^{-4}	36 ± 3	5×10^0
C124S	DMM	(1.21 ± 0.05) × 10^{-4}	3600 ± 300	3×10^{-2}
P240S[b]	DMM	11.7 ± 0.6	5900 ± 700	2×10^3
P240T	DMM	2.8 ± 0.2	4100 ± 800	7×10^2
WT	OA	(4.8 ± 0.2) × 10^{-1}	220 ± 30	2×10^3
Y44F	OA	< 1 x 10^{-5}	--------	-------
C124A	OA	(1.25 ± 0.01)×10^{-1}	142 ± 3	9×10^2
C124S	OA	(8.1 ± 0.2) ×10^{-2}	27 ± 3	3×10^3
P240S[b]	OA	(3.57 ± 0.08)×10^{-1}	1110 ± 70	3×10^2
P240T	OA	(8.2 ± 0.2) ×10^{-2}	1220 ± 3	7×10^1

[a]The same lack of catalytic activity was observed with 2R-methylmalate and 2R-ethyl-3S-methylmalate serving as substrate.
[b]Owing to insolubility of "native" recombinant DMML P240S, the His_6-tagged protein was prepared instead. It proved to be robust. The wild-type DMML was also prepared as the His_6-tagged protein and shown to possess the same level of catalytic activity as the wild-type DMML lacking the His_6-tag.

The counterpart to the PDP Ser257 in the *Botrytis cinerea* OAH is Ser260. The *B. cinerea* OAH activity (k_{cat}/K_m = 2 x 10^5 M^{-1} s^{-1}) (Han, Joosten et al. 2007) is 10-fold greater than that of PDP (k_{cat}/K_m = 2 x 10^4 M^{-1} s^{-1}) (Lu, Feng et al. 2005) and 100-fold greater than that of DMML (k_{cat}/K_m = 2 x 10^3 M^{-1} s^{-1}; Table 4.4). This higher OAH

activity has however been gained at the expense of the (2*R*,3*S*)-2-alkyl-3- methylmalate lyase activity, which is three orders of magnitude lower (k_{cat}/K_m = 2 x 10^2 M^{-1} s^{-1}) (Han, Joosten et al. 2007). In contrast, PDP (k_{cat}/K_m = 2 x 10^4 M^{-1} s^{-1} for (2*R*,3*S*)-2-ethyl-3-methylmalate) (Lu, Feng et al. 2005) and DMML (k_{cat}/K_m = 9 x 10^4 M^{-1} s^{-1} for (2*R*,3*S*)-dimethylmalate) (Table 4.3) have retained the (2*R*,3*S*)-2-alkyl-3-methylmalate lyase activity. The substitution of the PDP Ser257 with Pro reduces its OAH activity 10-fold and the (2*R*,3*S*)-2-ethyl-3-methylmalate lyase activity 2-fold (Joosten, Han et al. 2008). The substitution of the *B. cinerea* OAH Ser260 with Pro reduces its OAH activity 40-fold and increases the (2*R*,3*S*)-dimethylmalate lyase activity 15-fold (Joosten, Han et al. 2008). In the present study, the DMML P240S and P240T mutants were prepared and tested for DMML and OAH activities. Both activities drop in the two mutants, however the DMML activity is more greatly reduced (45 and 130-fold, respectively) than is the OAH activity (6 and 60-fold, respectively) (Table 4.4).

Together, these results suggest that the functional boundary between OAH and (2*R*,3*S*)-2-alkyl-3-methylmalate lyase activities among the fungal OAH and DMML and the plant PDP is not set solely by a single residue. Whereas the presence of OAH activity in the fungal OAH and plant PDP correlates with the presence of the Ser260 and Ser257, respectively, the fungal DMML also has substantial OAH activity despite the Pro at this position. Perhaps the Tyr44 in DMML compensates for the absence of a Ser at position 240, and indeed the fact that OAH also has a Tyr at this position may explain why its activity towards oxaloacetate is reduced only 40-fold in the S260P mutant. However, the OAH activity is only 10-fold lower in the PDP S257P mutant than in the wild-type enzyme yet the Tyr44 counterpart in PDP is Phe (as stated above, a water molecule may

substitute for the Tyr hydroxyl group). On the other hand, the presence of Pro (natural in the DMML and introduced by site directed mutagenesis in the OAH and PDP) rather than Ser at C(2)-substitutent binding pocket clearly enhances (2*R*,3*S*)-2-alkyl-3-methylmalate lyase activity.

Recognition of the C(3) substituent: The C(3) 3*S*-methyl group is important for the catalytic efficiency of DMML as is evidenced by the reduced k_{cat}/K_m (~72 fold) of 2*R*-methylmalate compared with (2*R*,3*S*)-dimethylmalate (and ~145-fold for 2*R*-ethylmalate compared with (2*R*,3*S*)-2-ethyl-3-methylmalate). The residues that delineate the C(3) methyl binding pocket are hydrophobic, including Ala48, Val216, Phe241, Pro240, and the partner subunit residues Pro272 and Phe276 (highlighted in green in Fig. 4.7). The C(3) methyl group appears to contribute to both binding affinity and orientation of the substrate with respect to the catalytic machinery. The absence of this methyl group is expected to diminish both types of contributions. Isocitrate is not a substrate of DMML, and 2-methylisocitrate is a very poor substrate, consistent with the electrostatic conflict that would be created by accommodating a C(3)CH_2COO- group in the hydrophobic pocket in place of the methyl group. Isocitrate and 2-methylisocitrate are also poor substrates for PDP owing to the clash between the C(3)CH_2COO- group and the hydrophobic side chains of Ile233, Leu258 and the benzyl ring of Tyr68 (Lu, Feng et al. 2005; Teplyakov, Liu et al. 2005).

4.4.8 Catalytic mechanism

Previously, we (Liu, Lu et al. 2005) and others (Britton, Langridge et al. 2000; Sharma, Sharma et al. 2000; Britton, Abeysinghe et al. 2001; Grimm, Evers et al. 2003) have proposed a catalytic mechanism for the lyase branch of the PEPM/ICL superfamily

that consists of deprotonation of the substrate C(2)OH group followed by protonation of the carbanion that develops on the C(3) as the C(2)-C(3) bond is cleaved (Liu, Lu et al. 2005). A proposed mechanism for DMML catalysis is shown in figure 4.9. The identity of the general base in ICL and MICL has not been demonstrated. We considered several possibilities: one of two Mg^{2+}-coordinated water molecules, each of which hydrogen bonded to a conserved carboxylic group (Asp59 and Glu116 in DMML); the invariant arginine (Arg161 in DMML); the Tyr-His pair discussed above (Tyr44 and His114 in DMML) (Liu, Lu et al. 2005). The present structure of 3,3-difluorooxaloacetate bound DMML (Fig. 4.6A) and the derived DMML/(2*R*,3*S*)-dimethylmalate model (Fig. 4.6B) exhibit the same groups as candidates for the role of general base. Because the $N^{\xi 1}$ and $N^{\xi 2}$ groups of Arg161 interact with both the C(3)COO- and the C(2)OH substrate groups, it is conceivable (although unusual) that a proton is transferred from the C(2)OH to the one nitrogen atom concomitantly with a proton that is transferred from the other nitrogen to the C(3)COO-. Alternatively, the Mg^{2+} water ligand bound to Asp59 and Lys122 is an attractive candidate for general base catalysis. Unfortunately, the site directed mutagenesis results do not allow us to assign the DMML general base. In addition to the replacement of Tyr44 by Phe that eliminated catalytic activity, we also replaced Asp59 with Ala and with Ser by site directed mutagenesis. The k_{cat}/K_m value for (2*R*,3*S*)-dimethylmalate cleavage is reduced 6-orders of magnitude in both mutants (Table 4.4). Similarly impaired activity was observed with Asp mutants in MICL and PDP (Grimek, Holden et al. 2003; Lu, Feng et al. 2005).

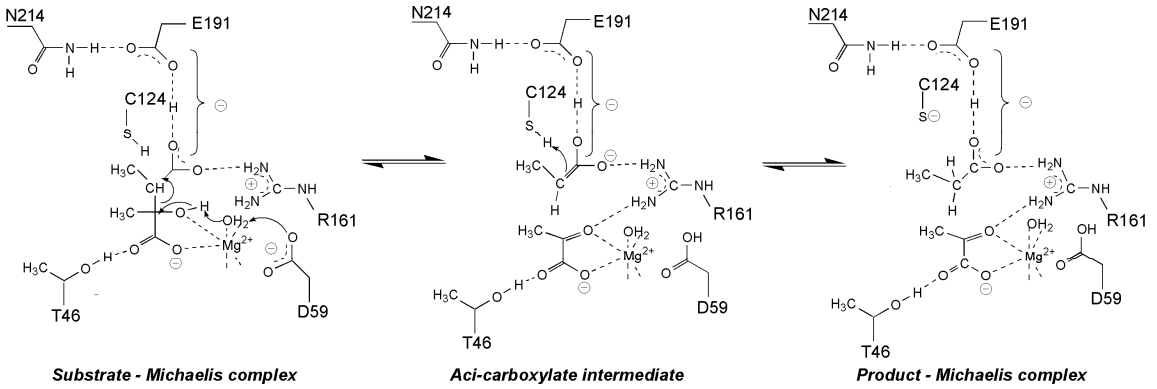


Figure 4.9: *Proposed catalytic mechanism of DMML.*

The candidates for general acid are the Cys124/Glu191 pair in DMML and their counter parts in ICL, MICL, and PDP. The Cys124 thiol group (assisted by the backbone amide NH of Gly125) is correctly orientated to protonate the C(3) upon C(2)-C(3) cleavage, whereas the Glu191 is correctly orientated to share a proton to the C(3)COO- (and is also engaged in hydrogen bond interaction with the side chains of Arg161 and Asn212 and the backbone amide NH of Cys124) (Fig. 4.6A). In the present study, the contribution that Cys124 makes to catalysis of (2*R*,3*S*)-dimethylmalate and oxaloacetate cleavage was probed by site directed mutagenesis. Because Cys124 does not bind either substrate the kinetic constant of interest is k_{cat}. The k_{cat} for oxaloacetate cleavage drops from 0.5 s^{-1} for the wild-type DMML to 0.1 s^{-1} for the C124A mutant and 0.08 for the C124S mutant (*viz*. ~5 fold) (Table 4.4). In contrast, the k_{cat} for (2*R*,3*S*)-dimethylmalate cleavage drops from 19 s^{-1} for the wild-type DMML to 2 x 10^{-4} s^{-1} for the C124A mutant and 1 x 10^{-4} s^{-1} for the C124S mutant (*viz*. 5 orders of magnitude) (Table 4.4). Why is Cys124 not required for oxaloacetate cleavage? One possible answer is that the pro-*S* C(2)OH (the binding OH) of the hydrated oxaloacetate acts as a surrogate proton donor.

Another intriguing question is why 3,3-difluorooxalacetate, an analogue of the oxaloacetate gem diol intermediate/transition state is not cleaved by DMML. From the outset, it would appear that the two fluoride atoms should stabilize the carbanion by delocalizing the negative charge and therefore the compound should be susceptible to cleavage. On the other hand, the prevalence of the C(2) gem diol form in solution suggests it as a more stable form than the ketone form of 3,3-difluorooxalacetate. This is reminiscent of fluorinated ketones and α-keto acid inhibitors of serine proteases. The electron withdrawing fluoride is thought to stabilize the diol form over the ketone, and these inhibitors add to the enzyme's catalytic serine residue to form a stable tetrahedral transition state analog (Imperiali and Abeles 1986; Stein, Strimpler et al. 1987; Brady, Wei et al. 1990).

4.4.9 DMML biological function

OAH homologs are ubiquitous in filamentous fungi. The genome of the acidogenic *A. niger* encodes four OAH family members. Genetic and biochemical evidence confirmed that An10g00820 encodes the protein responsible for oxalic acid formation (Han, Joosten et al. 2007). Another *A. niger* OAH paralog encoded by An07g08390 was identified by our studies as DMML. Current and previous analyses (Joosten, Han et al. 2008) indicate that locally conserved syntenic orthologs of DMML that share over 80% sequence identity are present in all *Aspergillus* species sequenced to date. Putative orthologs were also detected in more distantly related fungi.

In *Eubacterium barkeri* DMML is a key enzyme in nicotinate catabolism, catalyzing the last of nine enzymatic steps (Alhapel, Darley et al. 2006). The *E. barkeri* DMML shares 40% amino acid sequence identity with the *A. niger* DMML (recall the

57% sequence identity between the *A. niger* DMML and OAH). BLAST analysis (Altschul, Madden et al. 1997) of the *A. niger* genome for the presence of a similar pathway revealed the presence of some genes with sequence homology to *E. barkeri* nicotinate pathway enzymes. For instance, the small and large subunits of the *E. barkeri* 2,3-dimethylmalate dehydratase show 36% and 32% sequence identity with the two domains of homo aconitase, an enzyme involved in fungus-specific lysine biosynthesis encoded by An15g00350. Another example is the presence of a 3-methylitaconate isomerase homolog (39% identity). However, no homologs were identified for 2-methyleneglutarate mutase, 6-hydroxynicotinate reductase, 2-(hydroxymethyl)glutarate dehydratase, and enamidase. Thus from the genome sequence it is not evident that a nicotinate catabolic pathway similar to *E. barkeri* exist in *A. niger*. Moreover, the presence of the nicotinate pathway in *Aspergillus* has not been reported. Growth experiments revealed that when nicotinamide was used as a sole carbon source, *A. niger* remained viable but the growth was exceedingly poor.

An in house library of 140 *A. niger* Affymetrix DNA arrays (GEO: acc no GPL6758) was screened for conditions that support DMML expression. DMML was expressed when sorbitol was used as a sole carbon source and was repressed in the presence of fructose. CREA is a major regulatory protein that mediates carbon catabolite repression. In strain NW283 (a *creA d4* mutant strain relieved of carbon repression), the expression of DMML was derepressed. Taken together, the transcriptome analysis showed that unlike the OAH encoding gene (An10g00820), the DMML encoding gene (An07g08390) is under carbon catabolite repression. This accounts for why DMML does not compensate for OAH in the *A. niger* OAH gene knockout mutant, which does not

form oxalic acid. Therefore, the genetic and genomic data suggest that DMML plays a role in an evolutionarily conserved fungal pathway but presently this pathway is unknown.

4.5 Conclusions

For the *A. niger* mutant lacking the gene encoding OAH, the organism does not produce oxalate, yet its genome contains four more genes that exhibit sequence homology with the confirmed OAH, including one that shares 57% sequence identity with OAH. We show here that the latter gene encodes the enzyme 2,3-dimethylmalate lyase (DMML), and that this enzyme also catalyzes the hydrolysis of oxaloacetate albeit with 40-fold lower efficiency than it catalyzes the primary reaction. The OAH catalytic efficiency of DMML is 100-fold lower than the catalytic efficiency of an authentic OAH (Han, Joosten et al. 2007). The sequence-based phylogeny tree of the PEPM/ICL superfamily was constructed previously (Liu, Lu et al. 2005). Both DMML and OAH belong to a clade that also includes PDP and carboxyPEPM. Hence this is a clade of proteins of high sequence identities, typically 40% and above, that clusters together enzymes that share common catalytic machinery but perform multiple activities and exhibit broad substrate profiles. The 3D structures of representative members of the clade, PDP (Teplyakov, Liu et al. 2005), DMML (this work), OAH (unpublished results) and carboxyPEPM (unpublished results), are very similar, as is the conservation of the residues that define lyase activity. In contrast to this clade that contains multi-functional enzymes, two other clades encompassing primary metabolism enzymes, one defining MICL activity and the second ICL activity, contain highly specialized sequence families with specificity towards a single substrate. At the structural level, it appears that the

highly specialized enzymes evolved an optimized active site architecture that binds only one substrate productively, whereas the promiscuous enzymes contain features allowing accommodation of several substrates in a catalytically productive mode.

Accession numbers

The coordinates and structure factors have been deposited in the Protein Data Bank (entry codes 3FA3, 3FA4).

Chapter 5 Conclusion

The PEPM/ICL superfamily was chosen for study because of the interesting and diverse chemical reactions they perform (C-P and C-C bond formation and cleavage) even though they share similar overall fold and active site architecture, and exhibit high level of sequence identity. The biological and biochemical functions of some of these enzymes are already known, for others, the available structure/sequence information indicates that they might be incorrectly annotated in sequence databases and/or that they perform novel functions. The information gained from the structural studies was used to relate active site structure to catalysis and to identify markers, which were applied in the assignment of function to other family members. This work provides insights into the evolution of function from ancestral active site templates. Previously, the collaboration with Prof. Dunaway- Mariano's laboratory led to correction of the functional annotation of the petal death protein from CPEPM to a broad specificity lyase acting on (2*R*)-ethyl,(3*S*)-methylmalate, and oxaloacetate (Teplyakov, Liu et al. 2005). My own work led to the discovery of a novel biochemical reaction performed by *P. aeruginosa* PA4872, the decarboxylation of oxaloacetate and 3-methyl oxaloacetate (Narayanan, Niu et al. 2008) and to the discovery of the presence of (2*R*), (3*S*)-dimethylmalate lyase in the *A. niger* genome.

5.1 Structural insights

My studies of OAD and DMML confirmed that the common structural features of the PEPM/ICL superfamily, pertaining to the oligomerization state, overall fold, helix swapping, the presence of a capping loop, location of the active site, and the requirement

of a divalent metal ion holds true in these two new superfamily members. Yet, there were interesting features unique to each of these enzymes.

In OAD a novel mode of helix swapping occurs in which the C-terminus ensuing the swapped helix reverses direction compared with other structures (Fig. 3.2B). This alternative trace led to the formation of a channel extending from the active site to the solvent (Fig 3.5A), which may be utilized during catalysis in a manner that requires further investigation. DMML exhibits the common overall structural scaffold characteristic of the PEPM/ICL superfamily. In addition, DMML structures show that opening and closing of the active site capping loop of one molecule is coupled to the movement of the C-terminus of the partner molecule (Fig 4.5). Analogous movement has been observed in ICL (Sharma, Sharma et al. 2000) but not in PEPM. MICL exhibits a unique movement of its C-terminus.

5.2 Function fingerprints

In the PEPM/ICL superfamily, residues lining the active site perform a particular function: catalytic or substrate and metal cofactor binding. All members have the conserved Mg^{2+} coordinating residues. Presence of the conserved substrate coordinating residues such as the arginine and the N-terminus of a short helix enables the stabilization of the oxyanion intermediate. Presence of the lyase signature residues (Cys, Glu and Asn) infers the stabilization of an aci-carboxylate intermediate involved in the catalysis of all lyases, and OAH (along with a tetrahedral oxyanion intermediate). Among lyases the different enzymes are distinguished by the size and nature of residues surrounding the active site that define pockets for binding substrate groups. For example, the residues constituting the specificity triad dictates the nature of the C(2) substrate substituent.

Between the two highly specialized lyases of this superfamily (ICL and MICL), the presence of Ser315, Ser317, and Thr347 provides the fingerprint for distinguishing ICL from MICL, whereas the Arg241, Arg270, and Thr212 are the signature residues of MICL (Liu, Lu et al. 2005). Those enzymes that do not possess the above mentioned aci-carboxylate stabilizing residues stabilize an enolate oxyanion intermediate. Among these enzymes, a threonine or an asparagine at the position corresponding to the catalytic cysteine of the lyases on the gating loop suggests phosphonopyruvate hydrolase and phosphoenolpyruvate mutase, respectively. Among the PEPM/ICL superfamily members, CPEPM is an unusual candidate showing a lyase signature sequence yet performing a mutase activity via a pyruvyl enolate intermediate. More studies are required to understand the mechanism of this enzyme.

From the studies on DMML, we noticed that the counterpart of Pro240 in the OAH and PDP sequences is a serine. The results from our mutant studies (summarized in table 5.1), involving the P240S and P240T mutants of DMML, the S257P mutant of PDP and the S260P mutant of OAH (Joosten, Han et al. 2008) show that the presence of a proline enhances 2R-alkyl, 3S methylmalate lyase activity in OAH but not in PDP. On the other hand, the P240S mutant of DMML failed to show significant improvement of OAH activity. Thus, presence of a serine residue is not an obvious signature of OAH activity. In all these mutants, it is the K_m value that is affected the most indicating that Pro in DMML and Ser in OAH and PDP play an important role in substrate binding rather than catalysis. Moreover, the failure of these mutants to improve or impair specific functions underscores the complexity of evolutionary processes where function change is followed by further mutations that lead to subtle structural adjustments that optimize the

new function. Consequently, reversal of function cannot be engineered by a single mutation.

***Table 5.1**: Comparision of catalytic efficiency of the mutant and wild type proteins of DMML, OAH and PDP.*

	Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
DMML - WT	DMM	19.2 ± 0.5	220 ± 20	9×10^4
DMML - P240S	DMM	11.7 ± 0.6	5900 ± 700	2×10^3
OAH - WT	DMM	0.0293 ± 0.0007	140 ± 10	2×10^2
OAH - S260P	DMM	0.074 ± 0.002	22 ± 2	3×10^3
PDP - WT	DMM	8.4 ± 0.4	530 ± 30	2×10^4
PDP - S257P	DMM	8.3 ± 0.2	660 ± 40	1×10^4
DMML - WT	OA	$(4.8 \pm 0.2) \times 10^{-1}$	220 ± 30	2×10^3
DMML - P240S	OA	$(3.57 \pm 0.08) \times 10^{-1}$	1110 ± 70	3×10^2
OAH - WT	OA	17.4 ± 0.2	65 ± 3	2×10^5
OAH - S260P	OA	0.98 ± 0.03	220 ± 20	5×10^3
PDP - WT	OA	2.72 ± 0.06	130 ± 10	2×10^4
PDP - S257P	OA	4.1 ± 0.1	840 ± 50	5×10^3

Our studies of OAD revealed that at both the sequence and structure levels there are two residues unique to the OAD clade, yet the mutagenesis failed to confirm these residues as marker of OAD activity. The possibility that the capping loop may not undergo closure cannot be ignored because the loop lacks the conserved Lys residue. The presence of the channel leading to the active site is also not understood. The lack of loop Lys and its relatively hydrophobic nature may serve as the signature sequence of

decarboxylase activity but this awaits the structure determination of an enzyme/substrate complex.

5.3 Evolutionary perspective

In the phylogeny tree of the PEPM/ICL superfamily (Fig 1.5) the ICL and MICL enzymes are each clustered together. These two lyases have evolved to be highly specialized (Liu, Lu et al. 2005). Interestingly, within the highly specialized cluster of ICL, there is a second class of proteins called the ICL2's that possess MICL activity, suggesting a different route of evolution of some MICL's, closer to the authentic ICLs.

The structurally divergent transferases are well separated from the rest of the clusters. In this group, many characteristic features of the PEPM/ICL superfamily such as the presence of a capping loop, tetramer formation and metal coordination pattern differ. The grouping of the enzymes acting on a P-C bond, PEPM and PPHA, in the same cluster of the phylogeny tree underscores the possible evolution of these proteins from common ancestral active site architecture.

A cluster adjacent to the MICL cluster in the phylogeny tree consists of various enzymes that show diverse functional specialization despite their high sequence identity. This is an intriguing cluster that exhibits overlap of functions between its members, as exemplified by the fact that PDP, OAH and DMML all perform both OAH and lyase activities but with varied catalytic efficacies. The sequence differences in the active sites of these enzymes are small, hence the catalytic machinery of the lyases and OAH are related. The main difference is in the absolute requirement for the catalytic Cys for lyase activity whereas OAH activity can proceed without the Cys, albeit at lower efficacy. In general, these enzymes have broad substrate specificity unlike the specialized ICL and

MICL CarboxyPEPM remains an enigma because its mutase activity is extremely low in addition to its clustering with the lyases and OAH. My studies on a protein from this cluster revealed that the extremely close relative of OAH (~60% identity) is actually a DMML. Based on the present information this clade appears to be at a crossroad of function diversification and it may include additional enzymes with novel functions.

The phylogeny tree showed two distinct branches comprising proteins of unknown functions. From one of those branches we found the novel enzyme OAD from *P. aeruginosa* whose crystal structure was crucial to function assignment. The function of the second family remains unknown. Some of the sequences are annotated as "PEPM and related enzymes", some as "probable CPEPM", and others as hypothetical proteins. Recently, the structure of a family member was published containing an intrinsic α-ketoglutarate, but this structure does not provide clues about the function of the enzyme and its family relatives. The phylogeny analysis indicates that the functional diversity of the superfamily may be broader than is currently known, which means that there are numerous enzymes that catalyze novel C-C or P-C bond forming/cleaving reactions left to be discovered and analyzed.

5.4 Future directions

Additional X-ray structures of the PEPM/ICL superfamily members are needed to provide the basis for understanding the mechanisms of action at the atomic level and for understanding function diversification. We still do not understand the catalytic mechanism of OAD and expect that a structure of the enzyme/substrate complex will be informative. The biological role of the enzymes also needs further investigations. Within the DMML and PDP clade, the structures of OAH and CPEPM are yet to be determined,

which will hopefully clarify the underlying determinants of broad substrate specificity. There may be additional functions within this clade because *A. niger* contains four OAH paralogs (one of which is DMML). Finally, structure/function studies are required to discover the function of the remaining clade of “hypothetical proteins”. Once these goals are achieved, many more sequenced proteins will be annotated reliably.

Bibliography

Alhapel, A., D. J. Darley, et al. (2006). "Molecular and functional analysis of nicotinate catabolism in Eubacterium barkeri." Proc Natl Acad Sci U S A **103**(33): 12341-6.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Appel, R. D., A. Bairoch, et al. (1994). "A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server." Trends Biochem Sci **19**(6): 258-60.

Brady, K., A. Z. Wei, et al. (1990). "Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors." Biochemistry **29**(33): 7600-7.

Branden, C. I. (1980). "Relation between structure and function of alpha/beta-proteins." Q Rev Biophys **13**(3): 317-38.

Britton, K., S. Langridge, et al. (2000). "The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans." Structure **8**(4): 349-62.

Britton, K. L., I. S. Abeysinghe, et al. (2001). "The structure and domain organization of Escherichia coli isocitrate lyase." Acta Crystallogr D Biol Crystallogr **57**(Pt 9): 1209-18.

Brock, M., D. Darley, et al. (2001). "2-Methylisocitrate lyases from the bacterium Escherichia coli and the filamentous fungus Aspergillus nidulans: characterization and comparison of both enzymes." Eur J Biochem **268**(12): 3577-86.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." Nature **355**(6359): 472-5.

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

Chang, G. G. and L. Tong (2003). "Structure and function of malic enzymes, a new class of oxidative decarboxylases." Biochemistry **42**(44): 12721-33.

Chaudhuri, B. N., M. R. Sawaya, et al. (2003). "The crystal structure of the first enzyme in the pantothenate biosynthetic pathway, ketopantoate hydroxymethyltransferase, from M tuberculosis." Structure **11**(7): 753-64.

Chen, C. C., Y. Han, et al. (2006). "Structure and kinetics of phosphonopyruvate hydrolase from Variovorax sp. Pal2: new insight into the divergence of catalysis within the PEP mutase/isocitrate lyase superfamily." Biochemistry **45**(38): 11491-504.

Collaborative computational project (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Cowtan, K. (1994). DM: an automated procedure for phase improvement. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography.

Creighton, D. J. and I. A. Rose (1976). "Studies on the mechanism and stereochemical properties of the oxalacetate decarboxylase activity of pyruvate kinase." J Biol Chem **251**(1): 61-8.

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

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.

Erlick, H. A. (1992). PCR Technology Principles and Applications for DNA Amplication, W. H. Freeman and Co., New York.

Gibrat, J. F., T. Madej, et al. (1996). "Surprising similarities in structure comparison." Curr Opin Struct Biol **6**(3): 377-85.

Grimek, T. L., H. Holden, et al. (2003). "Residues C123 and D58 of the 2-methylisocitrate lyase (PrpB) enzyme of Salmonella enterica are essential for catalysis." J Bacteriol **185**(16): 4837-43.

Grimm, C., A. Evers, et al. (2003). "Crystal structure of 2-methylisocitrate lyase (PrpB) from Escherichia coli and modelling of its ligand bound active centre." J Mol Biol **328**(3): 609-21.

Han, Y., H. J. Joosten, et al. (2007). "Oxaloacetate hydrolase, the C-C bond lyase of oxalate secreting fungi." J Biol Chem **282**(13): 9581-90.

Hendrickson, W. A., J. L. Smith, et al. (1985). "Direct phase determination based on anomalous scattering." Methods Enzymol **115**: 41-55.

Horswill, A. R. and J. C. Escalante-Semerena (1999). "Salmonella typhimurium LT2 catabolizes propionate via the 2-methylcitric acid cycle." J Bacteriol **181**(18): 5615-23.

Huang, K., Z. Li, et al. (1999). "Helix swapping between two alpha/beta barrels: crystal structure of phosphoenolpyruvate mutase with bound Mg(2+)-oxalate." Structure **7**(5): 539-48.

Imperiali, B. and R. H. Abeles (1986). "Inhibition of serine proteases by peptidyl fluoromethyl ketones." Biochemistry **25**(13): 3760-7.

Jacobs, M. A., A. Alwood, et al. (2003). "Comprehensive transposon mutant library of Pseudomonas aeruginosa." Proc Natl Acad Sci U S A **100**(24): 14339-44.

Jones, C. E., J. M. Brook, et al. (1993). "Cloning and sequencing of the Escherichia coli panB gene, which encodes ketopantoate hydroxymethyltransferase, and overexpression of the enzyme." J Bacteriol **175**(7): 2125-30.

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

Joosten, H. J., Y. Han, et al. (2008). "Identification of fungal oxaloacetate hydrolyase within the isocitrate lyase/PEP mutase enzyme superfamily using a sequence marker-based method." Proteins **70**(1): 157-66.

Jursinic, S. B. and J. L. Robinson (1978). "The active site of rabbit muscle pyruvate kinase. Evidence for a site common to the oxalacetate decarboxylase and pyruvate kinase reactions." Biochim Biophys Acta **523**(2): 358-67.

Kanehisa, M., S. Goto, et al. (2006). "From genomics to chemical genomics: new developments in KEGG." Nucleic Acids Res **34**(Database issue): D354-7.

Kim, A., J. Kim, et al. (1998). "Isolation and characterization of the carbon-phosphorus bond-forming enzyme phosphoenolpyruvate mutase from the mollusk Mytilus edulis." J Biol Chem **273**(8): 4443-8.

Kleywegt, G. J. and T. A. Jones (1994). "Detection, delineation, measurement and display of cavities in macromolecular structures." Acta Crystallogr D Biol Crystallogr **50**(Pt 2): 178-85.

Kubala, G. and A. E. Martell (1981). "Nuclear Magnetic Resonance Investigation of the Spontaneous Decarboxylation of 2-Oxalopropionic Acid. 2. Species in Solution." J. Am. Chem. Soc. **103**: 7609-7615.

Kulakova, A. N., G. B. Wisdom, et al. (2003). "The purification and characterization of phosphonopyruvate hydrolase, a novel carbon-phosphorus bond cleavage enzyme from Variovorax sp Pal2." J Biol Chem **278**(26): 23426-31.

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

Laskowski, R. A., D. S. Moss, et al. (1993). "Main-chain bond lengths and bond angles in protein structures." J Mol Biol **231**(4): 1049-67.

Lenz, H., P. Wunderwald, et al. (1976). "Partial purification and some properties of oxalacetase from Aspergillus niger." Eur J Biochem **65**(1): 225-36.

Liao, C. J., K. H. Chin, et al. (2008). "Crystal structure of DFA0005 complexed with alpha-ketoglutarate: a novel member of the ICL/PEPM superfamily from alkali-tolerant Deinococcus ficus." Proteins **73**(2): 362-71.

Lill, U., P. Pirzer, et al. (1980). "Nicotinic acid metabolism enzymic preparation and absolute configuration of the substrate for 2,3-dimethylmalate lyase." Hoppe Seylers Z Physiol Chem **361**(6): 875-84.

Liu, S., Z. Lu, et al. (2004). "Conformational flexibility of PEP mutase." Biochemistry **43**(15): 4447-53.

Liu, S., Z. Lu, et al. (2005). "Crystal structures of 2-methylisocitrate lyase in complex with product and with isocitrate inhibitor provide insight into lyase substrate specificity, catalysis and evolution." Biochemistry **44**(8): 2949-62.

Liu, S., Z. Lu, et al. (2002). "Dissociative phosphoryl transfer in PEP mutase catalysis: structure of the enzyme/sulfopyruvate complex and kinetic properties of mutants." Biochemistry **41**(32): 10270-6.

Lu, Z., X. Feng, et al. (2005). "Diversity of function in the isocitrate lyase enzyme superfamily: the Dianthus caryophyllus petal death protein cleaves alpha-keto and alpha-hydroxycarboxylic acids." Biochemistry **44**(50): 16365-76.

M.D.Winn, M. N. I. a. G. N. M. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Cryst.(D57): 122-123.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J. Appl. Cryst.(40): 658-674.

McFadden, B. A., I. A. Rose, et al. (1972). "Production of pyruvate and succinate by action of isocitrate lyase on -methylisocitrate." Arch Biochem Biophys **148**(1): 84-8.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.

Narayanan, B. C., W. Niu, et al. (2008). "Structure and function of PA4872 from Pseudomonas aeruginosa, a novel class of oxaloacetate decarboxylase from the PEP mutase/isocitrate lyase superfamily." Biochemistry **47**(1): 167-82.

O'Leary, M. H. (1992). Catalytic Strategies in Enzymatic Carboxylation and Decarboxylation, in Mechanism of catalysis

D. S. Sigman, Ed., Academic Press, Inc., New York.**:** 239.

Otwinowski, Z. (1991). "Maximum Likelihood Refinement of Heavy Atom Parameters" In Isomorphous Replacement and Anomalous Scattering. Proc. Daresbury Study Weekend, SERC Daresbury Laboratory, Warrington, England.

Park, I. S. and R. P. Hausinger (1995). "Requirement of carbon dioxide for in vitro assembly of the urease nickel metallocenter." Science **267**(5201): 1156-8.

Park, S. H., B. G. Harris, et al. (1986). "pH dependence of kinetic parameters for oxalacetate decarboxylation and pyruvate reduction reactions catalyzed by malic enzyme." Biochemistry **25**(13): 3752-9.

Pearson, W. R. (1990). "Rapid and sensitive sequence comparison with FASTP and FASTA." Methods Enzymol **183**: 63-98.

Pearson, W. R. and D. J. Lipman (1988). "Improved tools for biological sequence comparison." Proc Natl Acad Sci U S A **85**(8): 2444-8.

Pel, H. J., J. H. de Winde, et al. (2007). "Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88." Nat Biotechnol **25**(2): 221-31.

Pflugrath, J. W. (1999). "The finer things in X-ray diffraction data collection." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1718-25.

Piccirilli, J. A., J. D. Rozzell, et al. (1987). "The stereochemistry of oxaloacetate decarboxylase: A stereochemical imperative?" J. Am. Chem. Soc. **109**: 8084-8085.

Pirzer, P., U. Lill, et al. (1979). "Nicotinic acid metabolism. 2,3-Dimethylmalate lyase." Hoppe Seylers Z Physiol Chem **360**(12): 1693-702.

Pollack, S. J., S. Freeman, et al. (1992). "Cloning, overexpression and mechanistic studies of carboxyphosphonoenolpyruvate mutase from Streptomyces hygroscopicus." Eur J Biochem **209**(2): 735-43.

Potterton, E., S. McNicholas, et al. (2002). "The CCP4 molecular-graphics project." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1955-7.

Potterton, L., S. McNicholas, et al. (2004). "Developments in the CCP4 molecular-graphics project." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2288-94.

Read, R. J. (1986). "Improved Fourier coefficients for maps using phases from partial structures with errors." Acta Crystallographica section A **42**: 140-149.

Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1373-82.

Rossmann, M. G. and D. M. Blow (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Cryst.(15): 24-31.

Ruijter, G. J., P. J. van de Vondervoort, et al. (1999). "Oxalic acid production by Aspergillus niger: an oxalate-non-producing mutant produces citric acid at pH 5 and in the presence of manganese." Microbiology **145 (Pt 9)**: 2569-76.

Schmitzberger, F., A. G. Smith, et al. (2003). "Comparative analysis of the Escherichia coli ketopantoate hydroxymethyltransferase crystal structure confirms that it is a member of the (betaalpha)8 phosphoenolpyruvate/pyruvate superfamily." J Bacteriol **185**(14): 4163-71.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.

Seidel, H. M., S. Freeman, et al. (1988). "Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond." Nature **335**(6189): 457-8.

Sender, P. D., M. G. Martin, et al. (2004). "Characterization of an oxaloacetate decarboxylase that belongs to the malic enzyme family." FEBS Lett **570**(1-3): 217-22.

Sharma, V., S. Sharma, et al. (2000). "Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis." Nat Struct Biol **7**(8): 663-8.

Simanshu, D. K., P. S. Satheshkumar, et al. (2003). "Crystal structure of Salmonella typhimurium 2-methylisocitrate lyase (PrpB) and its complex with pyruvate and Mg(2+)." Biochem Biophys Res Commun **311**(1): 193-201.

Stein, R. L., A. M. Strimpler, et al. (1987). "Mechanism for slow-binding inhibition of human leukocyte elastase by valine-derived benzoxazinones." Biochemistry **26**(13): 4126-30.

Studer, R., P. Dahinden, et al. (2007). "Crystal structure of the carboxyltransferase domain of the oxaloacetate decarboxylase Na+ pump from Vibrio cholerae." J Mol Biol **367**(2): 547-57.

Sugantino, M., R. Zheng, et al. (2003). "Mycobacterium tuberculosis ketopantoate hydroxymethyltransferase: tetrahydrofolate-independent hydroxymethyltransferase and enolization reactions with alpha-keto acids." Biochemistry **42**(1): 191-9.

Teplyakov, A., S. Liu, et al. (2005). "Crystal structure of the petal death protein from carnation flower." Biochemistry **44**(50): 16377-84.

Terwilliger, T. C. (2003). "Automated main-chain model building by template matching and iterative fragment extension." Acta Crystallogr D Biol Crystallogr **59**(Pt 1): 38-44.

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

van den Hombergh, J. P., P. J. van de Vondervoort, et al. (1995). "New protease mutants in Aspergillus niger result in strongly reduced in vitro degradation of target proteins; genetical and biochemical characterization of seven complementation groups." Curr Genet **28**(4): 299-308.

Vanni, P., E. Giachetti, et al. (1990). "Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme." Comp Biochem Physiol B **95**(3): 431-58.

von Delft, F., T. Inoue, et al. (2003). "Structure of E. coli ketopantoate hydroxymethyl transferase complexed with ketopantoate and Mg2+, solved by locating 160 selenomethionine sites." Structure **11**(8): 985-96.

Waldrop, G. L., B. F. Braxton, et al. (1994). "Secondary 18O and primary 13C isotope effects as a probe of transition-state structure for enzymatic decarboxylation of oxalacetate." Biochemistry **33**(17): 5262-7.