CRYSTAL STRUCTURE DETERMINATION OF METALLOPROTEINS:
PEPTIDE DEFORMYLASE, FIXL HEME DOMAIN, MONOMETHYLAMINE
METHYLTRANSFERASE, AND CARBON MONOXIDE DEHYDROGENASE

DISSertation

Presented in Partial Fulfillment of the Requirement for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Bing Hao, M.S.

*****

The Ohio State University
2002

Dissertation Committee:
Professor Michael K. Chan, Advisor
Professor Edward J. Behrman
Professor Ming-Daw Tsai

Approved by

___________________________
Advisor
Department of Biochemistry
This research was focused on the structure determination of four metalloproteins by x-ray crystallography. The first target was *E. coli* peptide deformylase that is responsible for deformylation of the N-terminus of nascent bacterial proteins and represents a potential drug target. We have determined the first crystal structures of formate- and inhibitor-bound deformylase complexes in different metal forms (Fe^{2+}, Co^{2+} and Zn^{2+}). The different formate-binding modes between the Zn and the other two metallated forms provide a possible explanation for the low activity of Zn enzyme as compared to Fe and Co enzymes. The inhibitor-bound structures reveal that the bound transition-state analog, (S)-2-O-(H-phosphonoxy)-L-caproyl-L-leucyl-p-nitroanilide (PCLNA), adopts an extended conformation and forms an interaction network with the protein. Based on these structures, a mechanism for deformylation is proposed and guidelines for the design of high-affinity deformylase inhibitors are suggested.

The second part of this research involved structural studies of the FixL heme domain from *Bradyrhizobium japonicum* (BjFixLH). FixL proteins are biological oxygen sensors that regulate nitrogen fixation gene expression in *Rhizobia*. In these proteins, the activity of the C-terminus kinase domain is regulated by the binding of O_2 and other strong-field ligands to the N-terminus heme domain. We have determined eight BjFixLH
structures including two unliganded and six ligand-bound forms. These structures reveal a novel heme-binding fold that has been conserved in PAS-domain sensor superfamily. Comparison of these structures has also revealed a heme-mediated conformational change that is distinct from that in classic globins. In BjFixLH, binding of O₂ to the heme results in the flattening of the heme plane, the rotation of a critical heme-pocket arginine, and the shift of a FG loop. We have proposed that this arginine plays a central role in ligand discrimination of BjFixLH.

The third project in this research involved the identification of an in-frame UAG-encoded amino acid in Methanosarcina barkeri monomethylamine methyltransferase (MtmB). We have determined the structures of MtmB in two crystal forms. Both structures suggest that the UAG-encoded residue is distinct from the other 21 natural amino acids. Instead it appears consistent with a lysine in amide-linkage to 4-substituted-pyrrole-5-carboxylate, which is named as L-pyrrolysine. Current data support the idea that pyrrolysine represents the 22nd genetically encoded amino acid. A model for how pyrrolysine participates in activating the methyl group of methylamine for transfer to corrinoid proteins is suggested.

The last target of this research was the α₂ε₂ carbon monoxide dehydrogenase (NiCODH) component of acetyl coenzyme-A decarbonylase synthase (ACDS) complex from Methanosarcina barkeri. This component catalyzes the reversible oxidation of CO to CO₂ at an unusual nickel-iron sulfur center. Our major effort has been devoted to the purification, crystallization, data collection, and phasing of this protein. So far, we have obtained two crystal forms and collected several MAD and MIR data sets with resolution
up to 2.2 Å. Five distinct iron centers have been identified in each αε monomer. The structure determination is in progress.
Dedicated to my parents

Weizhang Hao, and Kelin Wu
ACKNOWLEDGMENTS

This has been a long journey. For this reason, I would like to take a moment to thank the people who helped me finish this dissertation, as well as assisted me in studying and working in Dr. Michael Chan’s laboratory.

First, thanks to Dr. Chan, for giving me a chance to work in his laboratory as well as developing my research projects. Dr. Chan did a great job guiding me through my graduate studies and keeping me on track to finish my Ph.D. dissertation. Most importantly, he helped me to see what I have inside. I am grateful to his encouragement and enthusiasm in times of difficulty. I am honored to be his first student to get a Ph.D. degree.

Next, thanks to Dr. Weimin Gong who actually taught me crystallography and showed me the beauty of the macromolecular structures. It was a rewarding experience for me to work with him since I am now able to solve the structures independently.

A special thanks goes to my lab mate, Joseph Arndt. For five years, he helped me in many ways, both in the lab and on our “monthly” synchrotron trips. Also I greatly appreciate my other colleagues, Dr. Xuejun Zhong, Clara Isaza, Michelle Nauerth, Rinku Jain, and Patrick Kang, for their advice, assistance and happy times we spent together.
Also a huge thanks to Professors Dehua Pei, Marie A. Gilles-Gonzalez, Joseph Krzycki, and their group members for providing me protein used for structural analysis, without which, my research would have been impossible.

I am also grateful to Dr. Edward J. Behrman and Dr. Ming-Daw Tsai for serving on my dissertation committee.

In addition, the research done in this dissertation is supported by funds from the Ohio State University, the Petroleum Research Fund, the American Heart Association, and the NIH.

The last thanks goes to my parents and my brother for their continual and unlimited support and encouragement. Whatever happened, they were always there with me. I would never have been able to finish my Ph.D. degree without them. I hope this dissertation and degree make them be proud of me.
VITA

June 16, 1970………………………….. Born – Xian, China

1988 - 1992…………………………B.S. Microbiology, Xiamen University, China


1996 - present………………………Graduate Teaching and Research Associate, The Ohio State University

RELATED PUBLICATIONS


**FIELDS OF STUDY**

Major Field: Biochemistry  
Studies in X-ray Crystallography of Biological Macromolecules
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>VITA</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF PDB ID</td>
<td>xix</td>
</tr>
<tr>
<td><strong>CHAPTERS</strong></td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Metals in biology</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The highlight of this research</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Peptide deformylase: mechanism and inhibitor design</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 The sensory domain of BjFixL: a new mechanism for heme-driven signal transduction</td>
<td>3</td>
</tr>
<tr>
<td>1.2.3 Monomethylamine methyltransferase: identification of a novel UAG-encoded amino acid</td>
<td>4</td>
</tr>
<tr>
<td>1.2.4 Nickel carbon monoxide dehydrogenase: an unusual NiFeS cluster and the mechanism of CO oxidation</td>
<td>5</td>
</tr>
<tr>
<td>1.3 The basic procedure of x-ray crystallography</td>
<td>6</td>
</tr>
<tr>
<td>1.3.1 Crystallization</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Diffraction data collection</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Phasing</td>
<td>9</td>
</tr>
<tr>
<td>1.3.4 Model building and refinement</td>
<td>11</td>
</tr>
</tbody>
</table>
2. Crystal structures of the *Escherichia coli* peptide deformylase bound to formate and a transition-state analog ........................................ 18

2.1 Introduction .............................................................................................. 18
2.2 Materials and methods ........................................................................... 21
  2.2.1 Crystallization and data collection ....................................................... 21
  2.2.2 Structure determination and refinement .............................................. 23
2.3 Results and discussions .......................................................................... 24
  2.3.1 The protein fold .................................................................................. 24
  2.3.2 The metal active site .......................................................................... 25
  2.3.3 Binding of formate .............................................................................. 25
  2.3.4 Binding of PCLNA .............................................................................. 27
  2.3.5 Comparison to Thermolysin ................................................................. 32
  2.3.6 Comparison to other deformylase structures ...................................... 33
  2.3.7 A proposed catalytic mechanism for deformylation ............................ 34
  2.3.8 The origin of the specificity for N-formylated peptides ....................... 35
2.4 Conclusions .............................................................................................. 35

3. Crystal structures of the *Bradyrhizobium japonicum* FixL heme domain and ligand-bound complexes ....................................................... 53

3.1 Introduction .............................................................................................. 53
3.2 Materials and methods ........................................................................... 54
  3.2.1 Crystallization, sample preparation and data collection ......................... 54
  3.2.2 Structure determination and refinement .............................................. 56
3.3 Results and discussions .......................................................................... 58
  3.3.1 The overall heme-domain fold ............................................................. 58
  3.3.2 The PAS fold ...................................................................................... 59
  3.3.3 The heme pocket ................................................................................ 61
  3.3.4 Probing the role of the oxidation state of the heme iron on the protein conformation ................................................................. 62
  3.3.5 The conformation change in oxy-BjFixLH ........................................... 64
  3.3.6 Comparison of the Hydrogen bonding in BjFixLH and other O₂-binding heme proteins .................................................................................. 67
  3.3.7 Probing the role of spin-state and ligand sterics on the conformational change by the structures of NO- and CO-BjFixLH .......................... 68
  3.3.8 Investigating the role of electrostatic interaction in the conformational change from the structure of cyanomet-BjFixLH .......................... 72
  3.3.9 Evaluating the importance of the electronic interaction and a hydrogen bond in the heme distal pocket from the structures of imidazole- and Melm-BjFixLH .................................................. 73
  3.3.10 A proposed mechanism for allostery and ligand discrimination within BjFixLH .......................................................... 76
3.4 Conclusions .............................................................................................. 77
4. Crystal structure of a methanogen methyltransferase from *Methanosarcina barkeri*: identification of a new UAG-encoded amino acid ......................................................... 102

4.1 Introduction .......................................................................................................... 102
4.2 Materials and methods .......................................................................................... 105
  4.2.1 Crystallization and data collection.............................................................. 105
  4.2.2 Structure determination and refinement...................................................... 108
4.3 Results and discussions ........................................................................................ 110
  4.3.1 The protein fold of MtmB........................................................................... 110
  4.3.2 The UAG-encoded residue of MtmB.......................................................... 111
  4.3.3 A proposed mechanism for monomethylamine activation by pyrrolysine . 114
  4.3.4 Docking model of MtmB/MtmC ................................................................ 116
  4.3.5 Preliminary results for the crystallization and structural studies of the MtmBC complex ........................................................................................................ 117
4.4 Conclusion............................................................................................................ 119

5. Purification, crystallization, and structure determination of the nickel-containing carbon monoxide dehydrogenase (NiCODH) component of the acetyl coenzyme-A decarbonylase synthase (ACDS) complex from *Methanosarcina barkeri* ................. 139

5.1 Introduction .......................................................................................................... 139
5.2 Materials and methods .......................................................................................... 142
  5.2.1 Preparation of cell extracts and purification of the $\alpha_2\epsilon_2$ NiCODH component .................................................................................................................... 142
  5.2.2 Enzyme assays ............................................................................................ 144
  5.2.3 Crystallization and data collection.............................................................. 144
  5.2.4 The Patterson search and MAD phasing ..................................................... 146
5.3 Results and discussions ........................................................................................ 146
  5.3.1 Purification of the enzyme.......................................................................... 146
  5.3.2 The problems associated with crystallization and data collection.............. 147
  5.3.3 Identification of metal clusters and calculation of MAD phase ................. 148
  5.3.4 Comparison with other NiCODHs.............................................................. 149

BIBLIOGRAPHY........................................................................................................... 163
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Data processing and refinement statistics</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>Statistics for met-BjFixLH data collection, and phase determination.</td>
<td>79</td>
</tr>
<tr>
<td>3.2</td>
<td>Data processing and refinement statistics for deoxy-BjFixLH and ligand-bound complexes.</td>
<td>81</td>
</tr>
<tr>
<td>4.1</td>
<td>Statistics for data collection, and phase determination.</td>
<td>135</td>
</tr>
<tr>
<td>4.2</td>
<td>Refinement statistics.</td>
<td>138</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of data collected for crystal form 1 and 2 of the $\alpha_2\epsilon_2$ NiCODH core of the ACDS complex.</td>
<td>161</td>
</tr>
<tr>
<td>5.2</td>
<td>Data processing statistics for Kr and Xe derivatives of the $\alpha_2\epsilon_2$ NiCODH core of the ACDS complex.</td>
<td>162</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>(A) Selected elements important in biology. Naturally occurring elements are shown in color; those used as probes or drugs are depicted in gray. (B) Biological functions of selected metal ions.</td>
</tr>
<tr>
<td>1.2</td>
<td>Schematic diagram of protein crystallization methods: (A) the hanging drop vapor diffusion method; (B) the batch capillary method.</td>
</tr>
<tr>
<td>1.3</td>
<td>A typical single-crystal diffraction experiment</td>
</tr>
<tr>
<td>1.4</td>
<td>Flowchart for cryo cooling of crystals.</td>
</tr>
<tr>
<td>1.5</td>
<td>A plot of iron absorption edge.</td>
</tr>
<tr>
<td>2.1</td>
<td>Pathway for protein biosynthesis in prokaryotes.</td>
</tr>
<tr>
<td>2.2</td>
<td>Sequence alignment of deformylase proteins based on both primary and tertiary structures.</td>
</tr>
<tr>
<td>2.3</td>
<td>Crystals of <em>E. coli</em> peptide deformylase. (A) zinc enzyme bound to formate, and (B) cobalt enzyme bound to a transition-state analogue, PCLNA.</td>
</tr>
<tr>
<td>2.4</td>
<td>Structure of of the iron <em>E. coli</em> deformylase bound to formate. (A) Ribbon diagram of the secondary elements. (B) Electrostatic potential energy surface diagram of the protein in a similar orientation.</td>
</tr>
<tr>
<td>2.5</td>
<td>Ball-and-stick stereodiagrams of the metal sites of (A) iron deformylase bound to formate, and (B) zinc thermolysin bound to a phosphoramidates.</td>
</tr>
<tr>
<td>2.6</td>
<td>The assignment of formate. (A) $F_O-\tilde{F}_C$ omit density map at 6σ for active site of the cobalt deformylase grown in 3.6 M sodium formate. (B) $2F_O-\tilde{F}_C$ (blue, 3σ) and $F_O-\tilde{F}_C$ (violet, 6σ) density maps following incorporation of two water molecules. (C) $2F_O-\tilde{F}_C$ map at 3σ following incorporation of the formate molecule.</td>
</tr>
</tbody>
</table>
2.7 Ball-and-stick diagrams of the metal sites of formate-bound deformylases (A) iron form, (B) cobalt form, (C) zinc form, (D) superimposed iron (green), cobalt (pink), and zinc (purple) forms ................................................................. 44

2.8 Structure of the zinc *E. coli* deformylase-PCLNA complex. (A) Ribbon diagram of the secondary structural elements. (B) Electrostatic potential energy surface diagram of the protein and stick diagram of the PCLNA inhibitor in a similar orientation ................................................................. 45

2.9 Stereoview of the F₀-Fₑ electron density map of the active site of the zinc deformylase-PCLNA complex with the inhibitor omitted ................................................................. 46

2.10 (A) Protein-inhibitor hydrogen-bonding interactions within the zinc deformylase-PCLNA complex. (B) Interactions between inhibitor and protein ........................................ 47

2.11 (A) Stereoview of the active site of the zinc deformylase-PCLNA complex. (B) Superimposed stereoview diagram of the structures and interactions of phosphonate inhibitors bound to thermolysin (brown) and deformylase (blue). (C) Hypothetical model for the transition state involved in the deacylation of acetylated peptides ......................................................................................................... 48

2.12 Superposition of deformylase structures. (A) Overlap of native and formate-bound structures. (B) Overlap of inhibitor-bound structures ................................................ 50

2.13 Proposed mechanism of deformylation by *E. coli* peptide deformylase ............... 51

3.1 Crystals of met-BjFixLH ............................................................................................ 84

3.2 Harker sections of the Patterson maps of met-BjFixLH: (A) anomalous Patterson map using data from Nat3; (B) isomorphous Patterson map using data from Nat1 and Nat3 ......................................................................................................... 85

3.3 Ribbon diagrams of deoxy-BjFixLH ....................................................................... 86

3.4 Structure-based sequence alignment of BjFixL, RmFixL, PYP, HERG and LOV2 PAS domains .................................................................................................................. 87

3.5 The conserved PAS fold. (A) Overlap of PAS domain regions of PYP (yellow), HERG (blue), BjFixLH (red), and LOV2 (green). (B) *E. halophila* PYP colored according to the original PYP nomenclature .................................................. 88

3.6 (A) The structure of the heme pocket in deoxy-BjFixLH. The hydrogen bonds are shown as dashed lines; (B) Corresponding 2F₀-Fₑ electron density map (1σ) ..................... 89

3.7 Ribbon and ball-and-stick model diagrams of the BjFixLH heme-binding pocket for (A) met-BjFixLH, (B) deoxy-BjFixLH, (C) CO-BjFixLH, (D) NO-BjFixLH,
3.8 Ball-and-stick diagrams of three heme-binding pockets. Structures are shown for *Glycera dibranchiata* hemoglobin (PDB ID: 2HBG) (left), BjFixLH (center), and the NO transporter protein (PDB ID: 1NP1) (right). ............................................ 92

3.9 Overlap of the structures of deoxy-BjFixLH (blue) and met-BjFixLH (tan). ...... 93

3.10 Plots of RMSD following least square fitting of deoxy- and each of the six ligand-bound structures (CO-, NO-, cyanomet-, imidazole-, MeIm-, and oxy-BjFixLH) to met-BjFixLH: (A) The RMSD for each residue based on its main-chain atoms and (B) the overall RMSD for each structure based on all the atoms, the main-chain atoms, and the heme.............................................................................................. 94

3.11 (A) Ribbon diagrams of oxy-BjFixLH colored by secondary structure. (B) $F_O-F_C$ difference map (4.0σ) of the oxy-BjFixLH heme pocket with the dioxygen ligand omitted. ................................................................................................................. 96

3.12 Conformation differences in the FG loop (A) and in the heme pocket (B) between deoxy-BjFixLH (cyan and blue) and oxy-BjFixLH (red and tan). The structure of the heme pocket in oxy-BjFixLH (C) and the corresponding $2F_O-F_C$ electron density map (1σ). ................................................................................................................. 97

3.13 Structural comparison of the heme environment for (A), sperm whale oxy-myoglobin; and (B), oxy-BjFixLH. .................................................................................................. 98

3.14 (A) Ribbon diagram of CO-BjFixLH colored by secondary structure. Ball-and-stick models with electron density, for the BjFixLH heme binding pocket of (B) CO-BjFixLH, and (C) the co-crystallized cyanomet-BjFixLH (2$F_O-F_C$ density (blue, 1.0 σ) and $F_O-F_C$ omit density (red, 3.0 σ)). .................................................................................................................. 99

3.15 Proposed mechanism for the heme-driven conformational change of BjFixLH.101

4.1 Conserved pathways for methylotrophic growth............................................... 120

4.2 Dynamic light scattering measurement of the MtmBC complex......................... 121

4.3 Crystals of (A) the *M. barkeri* monomethylamine methyltransferase (MtmB), and (C) the complex of MtmB and its conjugated corrinoid protein MtmC (MtmBC). The SDS-polyacrylamide gel of the dissolved crystals of (B) MtmB and (D) MtmBC complex............................................................................................................. 122

4.4 Harker sections of the isomorphous (A) and anomalous (B) difference Patterson maps of the Hg derivative............................................................................................................. 123
4.5 Ribbon diagram of the MtmB hexamer. ................................................................. 124

4.6 Structure and folding of an MtmB subunit. (A) Ribbon diagram of one MtmB subunit. (B) Electrostatic potential energy surface diagram of MtmB in a similar orientation. ................................................................. 125

4.7 2F₀-Fᶜ electron density omit map of the UAG-encoded residue for crystals grown in (A) NaCl solution and (B) (NH₄)₂SO₄ solution.............................................. 126

4.8 Current fit of (4R, 5R)-4-substituted-pyrroline 5-carboxylate to the 2F₀-Fᶜ electron density of NaCl crystal form................................................................. 127

4.9 Electron density maps of the NaCl crystal form (A) 2F₀-Fᶜ omit map at 4.5σ for the 4-methyl-pyrroline carboxylate group of the UAG-encoded residue for crystals grown in NaCl solution with the two ring orientations used to fit it. The carbons of orientation 1 are colored in pink. In orientation 2 they are colored in tan. (B) Stick-diagram of proposed L-pyrolysine amino acid (orientation 1 in NaCl crystal form). (C) F₀-Fᶜ difference map at 4.5σ for NaCl crystals following incorporation of model for orientation 2 at 15% occupancy. (D) Fit of orientation 1 model to this density. (E) F₀-Fᶜ difference map at 4.5σ for NaCl crystals following incorporation of model for orientation 1 at 85% occupancy. (F) Fit of orientation 2 model to this density...................................................................... 128

4.10 Electron density maps of the (NH₄)₂SO₄ crystal form (A) 2F₀-Fᶜ omit map at 4.5σ for the 4-methyl-pyrroline carboxylate group of the UAG-encoded residue for crystals grown in high (NH₄)₂SO₄ solution with the two ring orientations used to fit it. The carbons of orientation 1 are colored in pink. In orientation 2 they are colored in tan. (B) Stick-diagram of proposed L-pyrolysine amino acid after amine addition (orientation 2 in (NH₄)₂SO₄ crystal form). (C) F₀-Fᶜ difference map at 4.5σ for (NH₄)₂SO₄ crystals following incorporation of model for orientation 2 at 60% occupancy. (D) Fit of orientation 1 model to this density with additional ammonium ion. (E) F₀-Fᶜ difference map at 4.5σ for (NH₄)₂SO₄ crystals following incorporation of model for orientation 1 at 60% occupancy. (F) Fit of orientation 2 model to this density with additional amine group added to the pyrroline ring. ..................................................................................................... 130

4.11 Stereoview of primary forms of the active site around the amber-encoded amino acid: (A) NaCl crystals; (B) (NH₄)₂SO₄ crystals, 40% occupancy orientation that is similar to NaCl crystals; (C) (NH₄)₂SO₄ crystals, 60% occupancy orientation with amine added to ring..................................................................................... 132

4.12 Hypothetical model for the role of the amber-encoded residue in catalysis...... 133

4.13 Overview of the docking model between MtmB and the cobalamin-binding domain of MetH (PDB ID: 1K7Y). (A) Molecular surface of MtmB and the stick
diagram of the cobalamin-binding domain of MetH. (B) Putative active site of the docking model.......................................................................................................................... 134

5.1 The quality of the $\alpha_2\varepsilon_2$ NiCODH component of the ACDS complex. (A) SDS-PAGE gel. (B) Dynamic light scattering measurement................................................. 152

5.2 Crystals of the $\alpha_2\varepsilon_2$ NiCODH core of the *M. barkeri* ACDS complex. .......... 153

5.3 Diffraction pattern of the crystal form 1 of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex................................................................................................................................. 154

5.4 Harker sections of the isomorphous difference Patterson maps of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex.......................................................................................... 155

5.5 The identified metal centers of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex. (A) The electron density of the metal clusters. (B) The corresponding distances between the clusters ............................................................................................................. 156

5.6 Comparison of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex with two known CODH structures. (A) The molecular surface of the $\alpha_2\varepsilon_2$ core. (B) Superimposed two known CODH structures.................................................................................................................. 157

5.7 The structures of C-clusters. (A) Proposed structure of C-cluster; (B) Dobbek’s model. (C) Drennan’s model.................................................................................................................. 158

5.8 Sequence alignment of the protein regions interacting with the metal clusters in NiCODH family.................................................................................................................. 159
## LIST OF PDB ID

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DFF</td>
<td>Zn deformylase</td>
</tr>
<tr>
<td>1BSK</td>
<td>Zn deformylase in complex with inhibitor (PCLNA)</td>
</tr>
<tr>
<td>1BSJ</td>
<td>Co deformylase in complex with inhibitor (PCLNA)</td>
</tr>
<tr>
<td>pending</td>
<td>Fe deformylase in complex with formate</td>
</tr>
<tr>
<td>pending</td>
<td>Co deformylase in complex with formate</td>
</tr>
<tr>
<td>pending</td>
<td>Zn deformylase in complex with formate</td>
</tr>
<tr>
<td>1DRM</td>
<td>met-BjfixLH</td>
</tr>
<tr>
<td>1LSW</td>
<td>deoxy-BjFixLH</td>
</tr>
<tr>
<td>1DP6</td>
<td>oxy-BjFixLH</td>
</tr>
<tr>
<td>1LSV</td>
<td>CO-BjFixLH</td>
</tr>
<tr>
<td>1DP8</td>
<td>NO-BjFixLH</td>
</tr>
<tr>
<td>1LT0</td>
<td>cyanomet-BjFixLH</td>
</tr>
<tr>
<td>1DP9</td>
<td>imidazole-BjFixLH</td>
</tr>
<tr>
<td>1LSX</td>
<td>MeIm-BjFixLH</td>
</tr>
<tr>
<td>1L2Q</td>
<td>MtmB at condition with (NH4)₂SO₄</td>
</tr>
<tr>
<td>1L2R</td>
<td>MtmB at condition with NaCl</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

1.1 Metals in biology

Metals are receiving ever-increasing attention due to their vital role in biology. Metalloproteins compose approximately one-third of known proteins, where they participate numerous biological processes (4). At least 14 metal ions have been found essential to the life (Figure 1.1) (5). These metals serve a wide variety of roles including: (1) Lewis acids in metallopeptidases, e.g. the zinc ion in carboxypeptidase A (6); (2) structure scaffolds, e.g. the zinc finger in transcriptional factor TFIIIA (7); (3) redox catalysts, e.g. the molybdenum and iron ions in nitrogenase (8, 9); (4) small molecule binding and transport, e.g. the iron in heme of hemoglobin (10); (5) electron transfer sites, e.g. the nickel-iron-sulfur clusters in carbon monoxide dehydrogenase (11); (6) a trigger for the electrical excitability of nerve cells, e.g. Na\(^+\) and K\(^+\) electrochemical potential gradients in the ion channel (11). Thus, to understand how metal ions in metalloproteins promote their given biological function is of fundamental interest.
1.2 The highlight of this research

The major goal of my research is to unravel the three-dimensional structure of metalloproteins and subsequently to explore the molecular basis of their functions based on these structures. In particular, my graduate research conducted in Dr. Chan’s laboratory includes peptide deformylase, a potential target for antibacterial chemotherapy; the FixL protein, a heme-based biological oxygen sensor; the MtmB, a methanogenic methyltransferase with an in-frame amber codon; and the α2ε2 component of the acetyl-CoA decarbonylase synthase (ACDS) complex, whose function in this complex is to catalyze the biological oxidation of CO with an unusual nickel-iron-sulfur cluster.

1.2.1 Peptide deformylase: mechanism and inhibitor design

Protein synthesis in bacteria involves the formylation and deformylation of the N-terminal methionine. As eukaryotic organisms differ in their protein biosynthetic mechanisms, peptide deformylase, the bacterial enzyme responsible for deformylation, represents a potential target for antibiotic studies. I have been involved in the crystallization, data collection, and structure determination and refinement of both the formate-and inhibitor-bound structures by MR methods in different metal forms (Fe$^{2+}$, Co$^{2+}$ and Zn$^{2+}$). A single metal ion is found in the active site bound tetrahedrally to three protein ligands, two histidines of a conserved HEXXH motif, and a cysteine, and a formate molecule serving as the forth ligand. The protein in complex with the transition-state analog, (S)-2-O-(H-phosphonyloxy)-L-caproyl-L-leucyl-p-nitroanilide (PCLNA), exhibits the same fold as the native enzyme. The PCLNA inhibitor adopts an extended
conformation and fits into a hydrophobic cavity located near the metal site. On the basis of these structures, a catalytic mechanism is proposed and guidelines for the design of high-affinity deformylase inhibitors are suggested.

1.2.2 The sensory domain of BjFixL: a new mechanism for heme-driven signal transduction

FixL proteins are biological oxygen sensors that regulate nitrogen fixation gene expression in *Rhizobia*. FixL’s oxygen-detecting domain is a heme-binding region that controls the activity of an attached histidine kinase. The FixL switch is regulated by binding of oxygen and other strong-field ligands. In this project, I have been involved in the crystallization, complex preparation, data collection, model building, and structure refinement of the *Bradyrhizobium japonicum* FixL heme domain (BjFixLH) (both ferric and ferrous forms) and a series of ligand-bound structures (oxy-, cyanomet-, imidazole-, NO-, CO-, and 1-methylimidazole-BjFixLH). These structures reveal a novel heme-binding fold that has been conserved in the PAS-domain sensory superfamily. These structures have also been used to develop a plausible mechanism of ligand discrimination and allostery for BjFixLH, which is distinct from the classical hemoglobin models. Unlike hemoglobin, the movement of the proximal histidine and its associated α-helix does not play a major role in FixL’s allosteric mechanism. Instead, binding of dioxygen to the FixL heme results in the flattening of the heme porphyrin, and release of a critical arginine to hydrogen-bond the bound dioxygen, followed by the shift of a critical loop (FG loop) away from the heme pocket. This FG loop shift presumably induces a global conformational change in the full-length protein that in turn inhibits the kinase activity.
The conserved arginine not only enhances the affinity of the bound dioxygen ligand in analogy to the distal histidine of classical globins, but also likely plays a critical role in ligand discrimination and allostery for BjFixL.

1.2.3 Monomethylamine methyltransferase: identification of a novel UAG-encoded amino acid

All genes encoding methanogenic methylamine methyltransferases contain an in-frame amber (UAG) codon that is read through during protein synthesis. To elucidate the identity of the UAG-encoded residues, we have determined crystal structure of the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB) at 1.55 Å resolution. In this project, I have been involved in all aspects of structure determination, including crystallization, data collection, isomorphous replacement phasing, model building and refinement. The MtmB structure reveals a homohexamer comprised of individual subunits with a TIM barrel fold that is reminiscent of other corrinoid-cofactor associated proteins. The UAG-encoded residue lies in the bottom of a deep cavity. The electron density for this residue is distinct from any of the 21 natural amino acids. Instead it appears consistent with a lysine in amide-linkage to 4-substituted-pyrroline-5-carboxylate, an amino acid that we suggest be named L-pyrrolysine. Combined with separate biochemical studies by Krzycki *et al.*, who have identified a downstream amber suppressor tRNA and a new aminoacyl-tRNA synthetase, these data support the case for L-pyrrolysine representing the 22nd genetically encoded amino acid. A model for how pyrrolysine participates in activating the methyl group of methylamine for transfer to corrinoid proteins is suggested.
1.2.4 Nickel carbon monoxide dehydrogenase: an unusual NiFeS cluster and the mechanism of CO oxidation

Species of Methanosarcina are methanogens capable of growing on acetate as the sole source of carbon and energy. In *M. barkeri*, the acetyl coenzyme-A decarbonylase synthase (ACDS) complex plays a central role in acetyl-CoA metabolism. According to the specific activities of the ACDS complex, it can be divided into three components. The $\alpha_2\varepsilon_2$ carbon monoxide dehydrogenase (NiCODH) component catalyzes the reversible oxidation of CO to CO$_2$ at an unusual nickel-iron sulfur center. In order to understand the mechanism of acetyl-CoA synthesis/degradation in methanogens, and to elucidate the structure of the nickel-iron sulfur cluster, we are working to determine the crystal structure of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex. My role in this project includes purification, crystallization, data collection, and MAD and MIR phasing. Due to the extreme oxygen sensitivity and the large size of this protein (214kDa), the purification, crystallization and structure determination have required extensive effort. So far, two crystal forms of this protein have been obtained and four sets of iron MAD data, and xenon and krypton derivative data have been collected with the resolution up to 2.2Å. Currently five distinct iron centers have been identified in each $\alpha\varepsilon$ monomer. The positions of these sites suggest a possible electron-transfer pathway. The structure determination is in progress.
1.3 The basic procedure of x-ray crystallography

The determination of three-dimensional structures of proteins, and protein complexes bound to substrates or inhibitors, has been proven to be a powerful approach on providing a basis for understanding numerous biological processes. My graduate studies have focused on using macromolecular crystallography to determine structures of proteins and protein complexes. The basic procedure has involved the crystallization of the target protein, collection of the diffraction data, phasing by various methods, model building, and refinement.

1.3.1 Crystallization

Getting high quality crystals from biological macromolecules has been a major obstacle in many structural investigations. To date, macromolecule crystallization is still dependent on empirical approaches rather than rational design though numerous theories have been developed to normalize it (12-14). The complexity of the crystallization stems from the fact that there are many parameters involved in the crystal nucleation and growth, such as pH, temperature, sample purity, and concentrations of specific organic additives and salts. A comprehensive search for the crystallization condition based on all the variables is almost impossible with normally limited sample source. Currently the most realistic and standard approach to find the initial crystallization condition is to use a sparse matrix sampling method to select reasonable numbers of trials from known crystallization conditions for macromolecules (15).

The initial crystallization conditions for the proteins used in my research were determined from two screen kits (up to 98 conditions) from Hampton Research (16). Two
crystallization techniques were used to obtain crystals: the hanging drop vapor diffusion method and the batch capillary method.

In the first method, 1-2 µl of protein solution is mixed with 1-2 µl crystallization reagent, and placed on a siliconized glass cover slide (Figure 1.2A). The cover slide is inverted and used to seal a reservoir containing a much larger volume (0.5 ~ 1.0 ml) of the crystallization reagent. The protein droplet is in vapor equilibration with the reservoir solution. Initially, the concentration of the crystallization reagent in the droplet is lower than that in the reservoir. As the concentration of the precipitants in the protein drop gradually increases, the protein begins to precipitate. Under the suitable condition, crystals form.

In the second batch method, a small volume of proteins and the crystallization solutions are placed into a cylindrical capillary without mixing (Figure 1.2B). Slow diffusion at the interface of the two solutions leads to slow precipitation, and under right conditions, crystals form. The batch method is particularly useful for anaerobic crystallization since two ends of the capillary can be sealed.

1.3.2 Diffraction data collection

X-ray data collection is the last experimental step in crystallography. It includes mounting the crystals in the x-ray beam, and measuring the intensity and pattern of the diffraction spots using an automated detector (Figure 1.3). Due to the difficulty of getting crystals, much effort is devoted toward maximizing the quantity and quality of the data that can be recorded from a single crystal. During the past ten years, many techniques have been developed that have enhanced the rate of structure determination by
crystallography. These include the greater availability of synchrotron radiation sources, widely used cryogenic technique, increased computational power, and improved softwares.

The electromagnetic radiation at synchrotron source is obtained by maintaining electrons in a circular orbit through magnetic forces (17, 18). The acceleration of the negatively charged ions required to maintain this circular orbit leads to the emitted radiation. The greatest benefits crystallographers can get from collecting data at synchrotron sources are high beam intensity and tunability. Synchrotron sources are hundreds of times more intense than conventional rotating anodes and highly collimated. The direct outcomes of this higher intensity are faster data collections and the ability to use much smaller crystals. Another key advantage of the synchrotron is that it emits a wide range of energies, allowing a beam of any wavelength to be produced. By using a tunable monochromator, one can select for the optimal energy of choice, enabling structure determination by multiple anomalous dispersion (MAD) method. Synchrotron data for my work were collected at the National Synchrotron Light Source (NSLS) at Brookhaven, NY, Advanced Photon Source (APS) at Argonne, IL, and Stanford Synchrotron Radiation Laboratory (SSRL) at Palo Alto, CA.

Another recent development in x-ray data collections is the routine use of cryo-temperature (around 100K) technique on both in-house facilities and synchrotron sources (19-21). The main reason for using this technique is to greatly reduce the x-ray-induced radiation damage to the protein crystals at low temperature. Currently, the simplest and most generally used method to freeze the crystal is the loop mounting method developed by Teng (22). It involves three steps (Figure 1.4). First, the crystal is transferred to a
cryo-solution containing both mother liquor and an anti-freeze agent, such as glycerol, oil, salt, or sucrose. Secondly, after a short soaking period, the crystal is lifted into a loop of thin nylon fiber (10-30µm) and plunged immediately into a cryogen such as liquid nitrogen, propane and carbon tetrafluoride. The use of the cryo-solution and rapid freezing hinders the formation of the ice and thus crystal damage. Finally, the frozen crystal can be either directly mounted to the goniometer using a cryotong or stored in a cryovial.

1.3.3 Phasing

The electron density of a crystallized macromolecule can be derived from the intensity and phases of each of the diffraction spots. During data collection, however, the detector only measures the intensity of the diffraction spots. All the phase information is lost. This is called the phase problem. Thus, a major challenge in macromolecular crystallography is determining these phases. Currently, there are three major methods that can be used to solve the phase problem for macromolecules (23).

The most common method is the multiple isomorphous replacement (MIR) method. In this method, the diffraction data from a crystal of a protein (native) are measured along a crystal to which an heavy atom, such as Hg, Pb, and Pt, has been added. Their diffraction intensity are compared and their difference detected. As these differences reflect the contribution of the heavy atom scattering, they can be used to derive the heavy-atom positions and phases, which in turn allows for the phase angles of the protein to be determined. The structures of *E. coli* peptide deformylase (Chapter 2) and monomethylamine methyltransferase (Chapter 4) were solved using this method.
Over the past several years, advances of synchrotron sources have facilitated the use of the multiple anomalous dispersion (MAD) method for phasing. This had a dramatic impact in macromolecular crystallography as evidenced by the increasing number of structures solved by this method. The MAD method is based on the concept of anomalous scattering. When the energy of the incident x-ray approaches the electronic absorption edge of an anomalous scatterer in the crystal, the scattering of that atom enlarges dramatically. By collecting data at the optimal wavelengths that maximize the real and imaginary anomalous differences, one is able to determine the location of the heavy-atom scatters and in turn obtain phase information needed to solve the structure. Generally, MAD data are collected at three different wavelengths, corresponding to the inflection point of the absorption edge ($\lambda_1$), the absorption maximum above the edge ($\lambda_2$), and a remote point at higher energy ($\lambda_3$). A plot for iron absorption edge is shown in Figure 1.5. Potential anomalous scatterers for use in MAD phasing include: metals inherent to protein (e.g. Fe, Mo, Zn), metal ions introduced as conventional heavy-atom derivatives (e.g. Hg, Sm, Pt), and selenium incorporated during protein biosynthesis (e.g. SeMet for Met). The structure of the BjFixL heme domain was solved using the Fe MAD method (Chapter 3). The positions of the iron-sulfur centers in $\alpha_2\varepsilon_2$ carbon monoxide dehydrogenase (CODH) have also been found from analyzing the Fe MAD data (Chapter 5).

Another common technique for solving the crystallographic phase problem is the molecular replacement (MR) method. This method can be used when a target protein has homology to a protein whose structure has already been determined. This search model could be a structural-related protein, or the same protein in different crystal form. In this
method, a rotation and translation function searches are carried out to find an orientation and position for the search model that can fit the diffraction data of the target protein. The structures of the peptide deformylase in complex with an inhibitor, PCLNA, were solved by the MR method (Chapter 2).

1.3.4 Model building and refinement

After the phase problem is solved, an electron density map can be calculated from the diffraction data and phases. Thus, the next step towards solving a structure is map interpretation and protein model building. Interpretation of an electron density map usually starts from fitting of secondary structure elements, such as α helix and β strands, and recognition of heavy-metal binding sites. And then, based on the shape, geometry and local and overall environments, the protein amino acid sequence is manually assigned into the electron density map. This process can be difficult and time-consuming since map quality and interpretability are highly sensitive to the initial phase. Following the first tracing, refinement of the initial model against the diffraction data is performed. Refinement is a process of adjusting the parameters of the model to obtain a best fit to the experimental data. These parameters typically include the positions, geometries, and temperature (B) factors of each atom in the model. The quality of the refinement is usually monitored and evaluated by the R-factor, which indicates the difference between the calculated model and observed data. Generally, multiple cycles of such model building and refinement are required until the R-factor is around 20%.
Finally, it is an incredible moment once the structure has determined. One can visually see the molecule and speculate how it works. This is the real beauty of crystallography.
Figure 1.1: (A) Selected elements important in biology. Naturally occurring elements are shown in color; those used as probes or drugs are depicted in gray. (B) Biological functions of selected metal ions (5).
Figure 1.2: Schematic diagram of protein crystallization methods (24): (A) the hanging drop vapor diffusion method; (B) the batch capillary method.
Figure 1.3: A typical single-crystal diffraction experiment (25).
Figure 1.4: Flowchart for cryo cooling of crystals (16).
Figure 1.5: A plot of iron absorption edge. The three wavelengths usually chosen in MAD phasing are labeled.
CHAPTER 2

Crystal structures of the *Escherichia coli* peptide deformylase bound to formate and a transition-state analog

2.1 Introduction

In all organisms, protein biosynthesis is initiated by the incorporation of a methionine residue (26). In prokaryotes and organelles of certain eukaryotes, however, an N-formylated methionine is used to begin the nascent peptide chain (27-29). In this process, a special initiator tRNA, tRNA\textsubscript{f}^\text{Met}, is charged with a methionine by methionyl-tRNA synthetase (Figure 2.1) (30), and then the methionine is formylated. The formylation is catalyzed by a methionine-tRNA\textsubscript{f}^\text{Met} formyltransferase that transfers the formyl group from 10-formyltetrahydrofolate to the charged methionine-tRNA\textsubscript{f}^\text{Met} (31-35). The majority of mature proteins in prokaryotes, however, do not retain their N-formyl group. And most of these proteins (about 60%) even lose their N-terminal methionine (36, 37). This observation suggests that a specific post- or co-translational modification might occur in the N-terminus of newly synthesized proteins. In fact, two steps are required to fulfill this task. Following translation initiation, first, the N-formyl group of the growing polypeptide is removed by an enzyme called peptide deformylase (PDF) (31, 38, 39). Then methionyl aminopeptidase will excise the remaining methionine to release the mature protein (Figure 2.1) (40, 41). Since the deformylation process is
necessary to obtain the final mature protein, peptide deformylase is thus essential for bacterial survival (42).

Since cytoplasmic protein synthesis in eukaryotic cells does not use N-formylmethionine in translation initiation, it has been widely believed that the formylation and deformylation process is a unique feature of bacterial cells. Indeed, peptide deformylase was thought to be absent in eukaryotes and archaea (42, 43). Based on these findings, deformylase has been suggested to be a potential antibacterial drug target since specific deformylase inhibitors could selectively block the growth of bacterial cells with little toxicity to a eukaryotic host (44-47). Consistent with these proposals, some compounds such as actinonin and hydroxamic acid derivatives have been proven to inhibit the deformylase activity and hinder bacterial cell growth (48-51). It should be noted, however, that several deformylase homologous genes have been identified in eukaryotic genomes recently, including protozoa, higher plants, animals, and humans (52-54). Several deforymlases from plants and protozoa have been characterized and have been shown to be localized in the organelles of the cell. The function of these proteins in mammals remains unknown.

Peptide deformylase is a novel metallopeptidase, which has been recently shown to utilize iron as the catalytic metal for amide hydrolysis (44). Purified deformylase has been reported to contain one metal atom per polypeptide (44). Biochemical or structural characterization of the iron form of the enzyme, however, has been hampered for three decades by its extreme instability. This instability has been finally overcome by the development of a purification procedure that limited exposure to oxygen (55, 56). Additionally, techniques have been developed to replace the iron with other metals such
as zinc (44, 57), nickel (58), or cobalt (59) by in vitro reconstitution or overexpression protocols. These alternate metal-forms are air stable and exhibit partial (Zn form) or nearly full catalytic activities (Ni and Co forms). One of the intriguing questions about deformylase is the origin of these activity differences, particularly since the tertiary structures of different metallated forms of deformylase have been shown to be nearly identical.

The sequences of deformylases are highly conserved in a wide variety of organisms (about 28-65% identity or similarity). They share three highly-conserved motifs: HEXXH, EGCLS, and G(I/V)G(L/I)AAXQ (Figure 2.2). Based on mutagenesis and structural studies, two histidines of the HEXXH motif and a cysteine of the EGCLS motif form the ligands for the metal ion (60, 61). The glutamine from the third motif, G(I/V)G(L/I)AAXQ, is also localized at the metal site and is thought to be involved in catalysis. One other feature of peptide deformylase is its high specificity for \( N \)-formylated substrates. It hydrolyzes acetylated peptides with rates more than four orders of magnitude lower than its formylated counterparts (62, 63).

In light of the key role that deformylase plays in prokaryotic protein synthesis and its potential application to antibacterial chemotherapy, structural studies were performed on \( E. \ coli \) deformylase. As part of this effort, we reported the first crystal structure of the zinc-containing \( E. \ coli \) deformylase (61).

Recently, we have solved three deformylase structures bound to Fe, Co, and Zn at high resolution. These structures reveal a bound formate at the metal active sites (Fe-, Co-, and Zn-formate). Importantly, binding of formate differs between the Zn and the
other two metallated forms. This observation may provide clues into the much lower activity of Zn-containing deformylase as compared to Fe and Co enzymes.

We have also determined the structures of the cobalt and zinc containing deformylase complexed with a transition-state analog, \((S)-2-O-(H-phosphonoxy)-L-caproyl-L-leucyl-p-nitroanilide\) (PCLNA) (64). These were the first structures of deformylase bound to an inhibitor, and the only one that mimics the physiological substrate. Taken together, these structures provide a solid basis for elucidating the mechanism of deformylation and facilitating the design of powerful therapeutic tools.

2.2 Materials and methods

2.2.1 Crystallization and data collection

Fe\(^{2+}\), Co\(^{2+}\), and Zn\(^{2+}\) containing \textit{E. coli} deformylases were overexpressed and purified in Dr. Pei’s laboratory according to published procedures (55, 57, 59). The initial crystallization conditions were determined via the hanging-drop method using a sparse matrix screen of 50 conditions at 4°C (65). Droplets containing 2 \(\mu\)L of protein solution [20 mg/ml protein, 20 mM NaH\(_2\)PO\(_4\) buffer (pH 7.0), 10 mM NaCl, 1.0 M (NH\(_4\))\(_2\)SO\(_4\)] and 2\(\mu\)L of precipitating solution (3.6 M sodium formate, with or without 0.1 M sodium acetate buffer, pH 4.6) were mixed on siliconized cover slides, inverted over a greased Linbro tray and sealed above a reservoir containing 0.5 ml of the precipitating solution. Small diamond-shaped crystals could be obtained at 4 °C after three days which grew to a size of 0.4 mm x 0.2 mm x 0.1 mm (Figure 2.3A). Tri(2-carboxyethyl)phosphine (TCEP) hydrochloride was used as a reducing agent to add to the crystallization drops of Fe\(^{2+}\) deformylase in a concentration of 1 mM.
The high-resolution data for deformylases were collected on three different metallated forms of these crystals at Beamline X4A at the National Synchrotron Light Source in Brookhaven National Laboratory (BNL), and Beamline BIOCARS 14C at the Advance Photon Source (APS) at Argonne National Laboratory (Table 2.1). Prior to data collection, the crystals were treated stepwise with mother liquor solutions containing increasing amounts of glycerol up to a final concentration of 40%(v/v). The crystals were then mounted on nylon loops and then cooled to 100K under a stream of nitrogen. Autoindexing and scaling were done with the HKL program suite (66). The crystals belong to space group P6122 with cell dimensions, \( a = b = 55.28 \text{ \AA}, c = 230.42 \text{ \AA}, \alpha = \beta = 90.0^\circ \) and \( \gamma = 120.0^\circ \).

The PCLNA inhibitor was chemically synthesized by Pei’s lab as reported (67). Crystals of both cobalt and zinc deformylase-PCLNA complexes were grown using hanging-drop vapor diffusion method at 4°C (Figure 2.3B). The reservoir solution contained 1 mL of 0.1 M Tris-HCl buffer (pH 8.5) and 1.9 M NH4H2PO4. Equal volumes of reservoir solution and protein solution (14.4 mg/mL protein, 20 mM NaH2PO4 buffer (pH 7.0), 10 mM NaCl, 0.86 M (NH4)2SO4, 1 mM inhibitor) were mixed. X-ray diffraction data were collected at Beamline X4A of BNL, using a RAXIS 100 detector (Table 2.1). The cobalt complex crystal diffracted to 2.0 Å resolution with cell dimensions of \( a = b = 98.38 \text{ \AA}, c = 109.37 \text{ \AA}, \alpha = \beta = 90.0^\circ \), and \( \gamma = 120.0^\circ \). The zinc complex crystal diffracted to 2.3 Å resolution with cell dimensions of \( a = b = 100.11 \text{ \AA}, c = 111.34 \text{ \AA}, \alpha = \beta = 90.0^\circ \), and \( \gamma = 120.0^\circ \). The space groups of both forms were either P6122 or P6522. Each unit cell contained one molecule per asymmetric unit.
2.2.2 Structure determination and refinement

The high-resolution structures (Fe-, Co-, and Zn-formate) were solved using the original deformylase structure (PDB entry: 1DFF) as the starting model. Model building and refinement were performed using the programs O (68) and CNS (69), respectively (Table 2.1). No sigma cutoff was used. 10% of the data was omitted for the Free-R factor calculation. The bound ligands and solvent molecules were modeled based on the resulting 2Fo-Fc and Fo-Fc electron density maps.

The two deformylase-PCLNA complex structures (Co- and Zn-PCLNA) were solved by the molecular replacement method using the program X-PLOR version 3.851 (70). The original deformylase structure (PDB entry: 1DFF) was used as the initial search model over the resolution range 12-3.0 Å. The correct solution was the highest peak in the rotation and translation search in the space group P6522. The refinement was performed with X-PLOR, using standard protocols and solvent correction (70) (Table 2.1). Modeling of the electron density was done using the program O (68). The location and conformation of the PCLNA inhibitor were determined based on 2Fo-Fc and Fo-Fc maps.

The quality of the model for each structure was evaluated using PROCHECK (71). All stereochemical parameters for each structure were within acceptable ranges with majority being better than acceptable. The programs, VERIFY-3D (72) and ERRAT (73), also gave acceptable values. One residue, Pro 9, adopts the cis conformation in all the structures. The root mean square deviations (RMSD) after least square fitting of the main-chain atoms of each of the complex structures (Co-formate, Zn-formate, Co-PCLNA and Zn-PCLNA) to Fe-formate structure were determined using the program
2.3 Results and discussions

2.3.1 The protein fold

E. coli peptide deformylase consists of a single domain of 168 amino acid residues. The overall folds of five structures (Fe-formate, Co-formate, Zn-formate, Co-PCLNA and Zn-PCLNA) are virtually identical. The RMSD between the main-chain atoms of the Fe-formate structure and other forms ranges from 0.17 to 0.43 Å. The Fe-formate complex structure is shown in Figure 2.4A. The protein fold consists of a mixed \( \alpha-\beta \) conformation, formed from three primary alpha helices, three beta sheets, and a potentially critical 3-10 helix. The secondary structure assignments determined using PROCHECK (71) and DSSP (77) are: \( \alpha \)-helix, (I) 11-14, (II) 25-40, (III) 124-136, (IV) 142-145, (V) 148-163; \( \beta \)-strands, (I) 45-47, (II) 57-60, (III) 70-81, (IV) 87-90, (V) 93-99, (VI) 105-111, (VII) 117-120; and 3-10 helix, 49-51. While two of the major \( \alpha \)-helices, helix II and helix V seem to play structural roles, the third major helix, helix III, appears to be involved in both structure and catalysis. In fact, the active site of the enzyme is found in a cleft formed by helix III, \( \beta \)-strands I, and IV (Figure 2.4B). The most conserved HEXXH motif is also located in this central helix and is apparently critical for metal coordination and substrate activation. A 3-10 helix (residues 49-51) is found near the active site. This short helix holds a conserved glutamine, Gln-50, which is important for catalysis. In each of the structures, the last two to four C-terminal amino acids were not observed.
2.3.2 The metal active site

All of our structures contain a mononuclear metal ion at the bottom of the active site cleft (Figure 2.4B). The active site is formed from three highly conserved motifs: G(I/V)G(L/I)AAXG (residues 43-52), EGCLS (residues 88-92), and HEXXH (residues 130-139). The protein ligands originate from two of these motifs, Cys 90 from motif E88-S92 and His 132 and His 136 from motif Q131-G139 (Figure 2.5A). This latter region contains the signature HEXXH sequence, which is found in many metallopeptidases including thermolysin (6, 78, 79). These structural features around the metal site are consistent with a role of the metal ion in catalysis (57). A formate or phosphonate serves as the remaining ligand.

A glutamate, Glu 133, on the HEXXH motif is positioned above the histidines of this motif in an orientation that is identical to the conserved glutamate in thermolysin (Figure 2.5B) (80, 81). Presumably it plays a similar role in the protonation and deprotonation of reaction intermediates during catalysis. Among the other non-coordinating residues near the metal site, two of the most important ones may be the conserved Gln 50 (from G(I/V)G(L/I)AAXG motif) and Leu 91 (from EGCLS motif). Both residues are hydrogen-bonded to the carbonyl oxygen of the formate, and hence are likely involved in the stabilizing and positioning the formyl group from cleavage.

2.3.3 Binding of formate

In our original Zn form structure, a water molecule was assigned as the fourth ligand to the metal based on the 2.9 Å electron densities (61). As part of our efforts to
elucidate the unusual metal dependence of peptide deformylase, we hence collected much higher resolution data for these different metallated forms of the enzyme, Fe form (1.85 Å), Co form (1.3 Å), and Zn form (1.85 Å). Using these high-resolution data, we were surprised to find extra density near the bound water that suggested the presence of a bound formate (Figure 2.6A). Bound formate gave a better fit to the density than either one or two waters, and is consistent with the high sodium formate concentration (3.6M) used in crystallization conditions. For comparison, the refined Fo-Fc and 2Fo-Fc density of Co-deformylase is shown as fit with two waters and with bound formate (Figure 2.6B and C).

The active sites of three formate-bound deformylase structures (Fe, Co and Zn) are shown in Figure 2.7. The two carboxyl-oxygen atoms of the formate molecule form a network of interactions with the surrounding residues. In Fe- and Co-formate complex structures, both the oxygen atoms coordinate to the metal ion in a bidentate fashion, resulting in the formation of a pseudo five-coordinated metal center (Figure 2.7A, and B). Hence, the oxygen atom close to Leu 91 residue forms the hydrogen bonds to the main-chain amide NH of Leu 91 and the side-chain amide NH2 of Gln 50, while the other oxygen atom hydrogen-bonds to the carboxylate side chain of Glu 133 and the carbonyl side chain of Gln 50. In Zn-formate complex structure, however, the formate-binding mode exhibits a potentially significant difference. Here, it appears that only one carboxyl-oxygen atom coordinates the zinc ion since one of the oxygen atom is much farther from the metal (3.06 Å verse 2.65 Å for first oxygen) (Figure 2.7 C).

Importantly, this observation may provide an explanation for the much lower activity of Zn deformylase relative to Fe and Co forms. The comparison on the Zn-, Fe-,
and Co-formate complex structures suggest that formation of a five-coordinate metal center is more facile for Fe and Co than Zn. The ability to easily form a five-coordinate intermediate may be required for optimal turnover of the enzyme (61, 82). This may be the case here in light of the small steric cone angle of the formate ion. An interesting question that these results engender is why the zinc adopts the different binding mode. One explanation could be that zinc requires a more regular geometry than the other metals (Fe, Co, and Ni) (83).

2.3.4 Binding of PCLNA

While the formate-bound deformylase complexes provide a picture for how the product binds to the active site, more interesting would be the interactions of a bound transition state inhibitor. Such a structure would help to understand the mechanism of deformylation and to elucidate the interactions that contribute to the binding affinity of inhibitors. Such information could be critical for the design of more potent inhibitors. The PCLNA inhibitor, (S)-2-O-(H-phosphonoxy)-L-caproyl-L-leucyl-p-nitroanilide, was developed as a non-hydrolyzed transition-state analogue of the potent substrate, f-Met-Leu-p-nitroanilide (63, 67). It incorporates a H-phosphonate group to mimic the tetrahedral intermediate formed during formyl hydrolysis. L-caproyl was utilized in place of L-methionine to facilitate its synthesis. PCLNA acts as a competitive inhibitor of both the zinc and iron forms of PDF with $K_i$ values of 76 µM and 37 µM, respectively (67).

The structures of both the Co- and Zn-PCLNA complexes are virtually identical. The overall fold and electrostatic potential energy surface diagram of the zinc form is shown in Figure 2.8. The location and orientation of the PCLNA inhibitor were
determined from the Fo-Fc and 2Fo-Fc maps. It adopts an extended conformation similar to a β-strand and inserts into a cleft formed by the active-site helix (125-136) on the bottom, the β-strand I (43-47) on one side, the β-strand IV (87-91) on the other side, and the 3-10 helix at the back (Figure 2.9).

As PCLNA is a peptide-based transition state analog, our complex structures allow for an understanding of how deformylase recognizes nascent peptides, its physiological substrate, and mediates the hydrolysis of the N-formyl peptide bond. One important global feature relevant to the binding of the PCLNA inhibitor may be our observation that the PCLNA peptide, together with β strands IV and V from the protein, forms an antiparallel β sheet. Three hydrogen bonds are formed between the main-chain atoms of the inhibitor and the β strand IV of the protein (Figure 2.10A). The formation of this sheet involves cooperative interactions and probably contributes significant stability to the peptide binding affinity. These considerations suggest a potential role for these specific secondary structural elements within the deformylase proteins.

Another factor may be the ability of the PCLNA inhibitor to fill the hydrophobic cavity (Figure 2.8B). Enhanced protein stability has been shown to be associated with removal of cavities, resulting from the reduction of the unfavorable surface energy and the increase of the favorable hydrophobic packing contacts. Calculations of the solvent accessible surface areas of the deformylase protein only (8,887 Å²), the PCLNA inhibitor (692 Å²), and the deformylase-PCLNA complex (8,720 Å²) using program CNS reveal that the binding of the PCLNA inhibitor to deformylase results in a lower molecular surface area for the protein in general, burying a total of 859 Å² of surface area (69, 84).
These results support the potential importance of these effects on the PCLNA binding affinity.

There are, of course, individual interactions between residues of the PCLNA inhibitor and the protein that are important for binding (Figure 2.10B). First, the \( p \)-nitroanilide (\( p \)-NA) group at the C-terminal end of the PCLNA inhibitor (S\(_3^\prime\), farthest from the metal ion) has been shown to be critical for binding \((63)\). For example, f-Met-Leu-\( p \)-nitroanilide is a potent substrate of the \( E. coli \) iron deformylase \((k_{cat} \text{ of } 38 \text{ s}^{-1}, \text{ a } K_M \text{ of } 20 \text{ M}, \text{ and a } k_{cat}/K_M \text{ of } 1.9 \times 10^6 \text{ M}^{-1}\text{s}^{-1})\). Removal of the C-terminal \( p \)-NA group reduces the enzyme’s affinity to f-Met-Leu-NH\(_2\) by ~40 fold \((k_{cat} = 194 \text{ s}^{-1}, K_M = 840 \text{ M}, k_{cat}/K_M = 2.3 \times 10^5 \text{ M}^{-1}\text{s}^{-1})\). However, the mapping of the interactions of the \( p \)-NA group with the PDF protein in the present study is complicated by the fact that the \( p \)-NA also interacts with a symmetry-related molecule generated by a crystallographic two-fold located near the active site. As a result, the \( p \)-NA of the PCLNA inhibitor interacts with residues from both the protein to which it is bound as well as residues from the symmetry related protein. One face of the hydrophobic pocket around the bound \( p \)-NA group is formed by the side chains of three protein residues, Ile 44, Ile 86, and Leu 125, the backbone carbonyl of Glu 87, and the caproyl side-chain of the PCLNA inhibitor itself. The other side of the \( p \)-NA group is in close contact with the backbone and/or side chains of Gly’ 124, Leu’ 125, and Ile’ 128 of the symmetry related protein. While there are undoubtedly additional interactions that stabilize the protein-protein interface between the two molecules, no other experiments that we have performed to date support the presence of a dimer in solution.
The second residue of the PCLNA inhibitor, L-leucine (S$_2'$), is in van der Waals contact with Leu 91 and the main-chain carbonyl groups of residues Glu 42 and Gly 43 on one side, while the other side is exposed to solvent. The backbone amide and carbonyl of the leucyl residue are also hydrogen-bonded to the corresponding carbonyl and amide of Gly 89. These interactions are to be expected when the peptide is involved in the formation of an antiparallel β-sheet. Since the side chain of this group is directed towards an open cavity, as shown in Figure 2.8B, the enzyme should be able to accommodate a number of different amino acids at this position.

The N-terminal L-caproyl group (S$_1'$, closest to the metal ion), which mimics the methionine amino acid, sits in a pocket generated by residues Gly 43, Ile 44, Gly 45, Glu 88, Cys 129, His 132, and Glu 133, as well as the p-NA group of the PCLNA inhibitor itself. The carbonyl of the caproyl group is hydrogen-bonded to the main-chain amide nitrogen of Ile 44. The most interesting feature, though, may be the interaction of the caproyl group with the p-NA group. As a result of the β-strand conformation of the PCLNA inhibitor, the p-NA group acts as a lid, burying the caproyl side-chain in a hydrophobic cleft of the protein. These intimate hydrophobic interactions between the protein and the inhibitor presumably make a significant contribution to the binding affinity of this inhibitor, and may provide a potential explanation for the enzyme’s strong preference for a methionine residue at the N-terminus of peptide substrates (57).

Finally, the structures of the PCLNA-deformylase complex provide important insight into the interactions of the H-phosphonate group with the protein. These interactions are important since the H-phosphonate is a transition-state analog of the tetrahedral intermediate formed during deformylation. As expected, this group binds to
the zinc metal via one of its phosphonate oxygens. This phosphonate oxygen is also hydrogen-bonded to Glu 133, consistent with the putative role of Glu 133 as a proton shuttle, receiving the proton from the metal bound water and then donating it to the amide NH to facilitate bond cleavage (Figure 2.11A). The carbonyl group of Gln 50 is within hydrogen-bonding distance (3.1 Å) of this phosphonate oxygen as well.

The second phosphonate oxygen of the inhibitor, which mimics the carbonyl oxygen of the formyl group in the enzyme-substrate complex, forms hydrogen bonds to the main-chain amide NH of Leu 91 and the side-chain amide NH$_2$ of Gln 50. The third phosphonate oxygen, whose position would correspond to the N-terminal methionine amide nitrogen of nascent peptides, is not involved in hydrogen bonding in the deformylase-PCLNA complex. However, it is within hydrogen-bonding distance (3.1 Å) of the carbonyl of Gly 45. Presumably in the physiological reaction, the formylated N-terminal amide NH of the nascent peptide hydrogen bonds to this carbonyl group.

In total, in addition to hydrophobic and metal binding interactions, the deformylase-PCLNA complex structure reveals seven hydrogen bonds that could account for the structure of the bound inhibitor in the enzyme and its binding affinity (Figure 2.10A). An important issue in the design of antibiotics is, of course, the generality of these interactions over different organisms. Figure 2.2 summarized the results of sequence analysis of the residues that must interact with the PCLNA inhibitor (within 4.0 Å) over deformylase proteins from a wide variety of organisms given the mode of inhibition inferred from the present study. The important result is that these residues are highly conserved among all deformylase proteins, including those with low overall sequence homology. This outcome is not unexpected in light of the generality of the
deformylation reaction, and it suggests that drugs targeting deformylase will have broad applicability.

### 2.3.5 Comparison to Thermolysin

While the tertiary and secondary structure of peptide deformylase differs from other zinc metalloproteases (85), the active site environment has strong homology to other zinc metalloenzyme active sites. Most notable is its similarity to proteins of the thermolysin family, all of which contain the characteristic HEXXH motif (78, 80). The superposition of the metal sites of deformylase and thermolysin in complex with the inhibitors is shown in Figure 2.11B. Both structures contain a tetrahedral metal ion coordinated by three protein ligands and a carboxylic oxygen from the inhibitor molecule. While deformylase contains a cysteine residue and thermolysin contains a glutamate, the remaining two protein ligands are histidine and the overall positioning of the coordinated ligands is similar. Moreover, as expected, the conserved Glu from the HEXXH motif is in a similar location, but perhaps somewhat surprising is the number of other hydrogen-bonding interactions that are also conserved, although using different residues. In thermolysin a tyrosine and histidine (Tyr 157 and His 231) provide important hydrogen bonding interactions to the phosphate oxygen that binds to the metal iron. In deformylase, the similar interactions are provided by the side chain of a glutamine and the backbone amide nitrogen of a leucine (Glu 50 and Leu 91). Also conserved in both structures is a beta sheet region formed from beta strands I, II, and III in deformylase and a similar motif in thermolysin. This beta sheet motif appears to be a conserved feature of all HEXXH metalloproteases.
While the inhibitors in both complex structures contain a bound phosphonate group, one surprising difference is the orientation of the phosphonate. In the deformylase-PCLNA complex, the phosphonate oxygen, which does not bond to the metal, forms a hydrogen bond with the main-chain amide NH of Leu 91. For all other related metalloproteases bound to phosphonate inhibitors that we have analyzed, this phosphonate oxygen hydrogen-bonds the catalytic glutamate (Glu 143 in thermolysin, Glu 133 in \textit{E. coli} deformylase). One important consequence of this different binding orientation is the fact that in this orientation deformylase could only be cleaved by a metal hydroxide mechanism. This is since the Leu 91 amide NH can only act as a general acid and not as a base, the nucleophilic hydroxide must come from the metal.

\subsection*{2.3.6 Comparison to other deformylase structures}

Recently, Wagner and his coworkers determined a series structures of \textit{E. coli} deformylase in the forms of ligand-free (Ni\textsuperscript{2+}, Zn\textsuperscript{2+}), bound to polyethylene glycol (Fe-, Ni-, and Zn-PEG), and in complex with the product Met-Ala-Ser (Ni-, and Zn-MAS), respectively (58, 82). Moreover, structures of deformylases from bacteria, \textit{Staphylococcus aureus}, \textit{Bacillus stearothermophilus}, \textit{Pseudomonas aeruginosa}, and protozoa, \textit{Plasmodium falciparum}, have also been determined (86-88). Furthermore, the structures of deformylases bound to the antibiotic agent actinonin from four different bacteria were also reported by Guilloteau \textit{et al}. (86).

The overall folds and active sites of these deformylase structures are nearly identical to \textit{E. coli} deformylase (Figure 2.12A). The position and conformation of the bound inhibitors and products in these structures are also similar to the PCLNA in our \textit{E.}}
coli deformylase structures (Figure 2.12B). The most significant difference of these structures from different organisms is the conformation around the C-terminus. For instance, the helix V in *E. coli* deformylase is replaced by a β strand in deformylase from *Staphylococcus aureus* (86, 87). This conformation variation likely arises from having to accommodate insertions into the sequence of *S. aureus* deformylase. As it is far from the active site, it is unlikely to be relevant to the catalysis.

### 2.3.7 A proposed catalytic mechanism for deformylation

Based on the available structural information, a plausible mechanism for deformylation is depicted in Figure 2.13. This proposal is similar to the mechanism suggested for thermolysin (80). In this model, the metal initially serves as a Lewis acid to stabilize the bound water or hydroxide molecule, which is also hydrogen-bonded to Glu 133 (as a general base) (82, 88). When the formylated peptide binds to the active site, the nucleophile, the metal-bound hydroxide, attacks the carbonyl carbon of the formyl group to form a tetrahedral intermediate. Here, the carbonyl oxygen of the formyl group is stabilized by a pair of hydrogen bonds from both the side-chain amide of Gln 50 and the main-chain amide of Leu 91. Based on the high-resolution formate-bound deformylase complex structures, the metal in the transition state is presumably pentacoordinated with both oxygen atoms of the tetrahedral intermediate. In addition, the protonated Glu 133 (as a general acid) donates the hydrogen to the amide nitrogen of the substrate to make it a better leaving group, which hence facilitates the cleavage of the facile amide bond. Consistent with this mechanism, mutations at Glu 133 (to Ala, Asp, or Gln) eliminate the
catalytic activity of the enzyme (62). It is important to note that other variations of this mechanism are possible.

2.3.8 The origin of the specificity for N-formylated peptides

Biochemical studies have revealed that while deformylase efficiently deformylates peptides containing an N-terminal formylated Met, it is much less active towards the corresponding $N$-acetyl-Met-peptides (63, 89). These observations have been one of the intriguing issues regarding this enzyme. Based on the structures of the deformylase-PCLNA complexes, we have modeled the transition state for deacylation of $N$-acetylated peptides, by replacing the proton on the H-phosphonate group with a methyl group (Figure 2.11C). This analysis revealed that the modeled methyl group forms a close interaction (2.2 Å) with the carbonyl oxygen of Gly 45, one of the potentially important residues involved in the catalytic mechanism. This steric interaction would hinder the deacylation of acetylated substrates. Thus, *E. coli* peptide deformylase may have evolved to use the altered orientation of the transition state to achieve its catalytic selectivity.

2.4 Conclusions

In this project, we have determined the x-ray structures of *E. coli* peptide deformylase in complex with either formate or a transition-state analogue, PCLNA, in different metal-bound forms. These structures have provided the first insights into the overall fold conserved in all deformylases examined to date and allow us to propose a mechanism that involves a metal-bound hydroxide in the deformylation reaction. In
addition, the difference in the binding mode of formate to the active sites of deformylase sheds light on how the metals affect the enzymatic activity of deformylase. Furthermore, the deformylase-PCLNA complex structures reveal the important protein-substrate interactions and offer certain guidelines for the design of high-affinity deformylase inhibitors. Finally, based on the orientation of the PCLNA inhibitor in the active site, steric hindrance is suggested to be the origin of specificity of deformylase for formylated peptides over acetylated substrates.
Figure 2.1: Pathway for protein biosynthesis in prokaryotes.
Figure 2.2: Sequence alignment of deformylase proteins based on both primary and tertiary structures. The secondary structural elements of *E. coli* deformylase are drawn above the sequences (helices, cylinder; strands, arrows; loops, lines), and labeled as in text. Three highly reserved signature motifs are boxed. The residues that interact with the transition state inhibitor, PCLNA, are colored by the region of the inhibitor they interact with: H-phosphonate (violet), L-caproyl (blue), L-leucine (green), and nitrophenylaniline (yellow). The color is alternated for the residues that interact with more than one region. The PDB ID for the deformylase from, *P. stearothermophilus*, *P. aeruginosa*, *S. aureus*, and *P. falciparum* are 1LQY, 1LRY, 1LQW, and 1JYM, respectively. Other sequences were obtained from the EBI databases (http://www.ebi.ac.uk/dbases/topdata.html).
Figure 2.2
Figure 2.3: Crystals of *E. coli* peptide deformylase. (A) zinc enzyme bound to formate, and (B) cobalt enzyme bound to a transition-state analogue, PCLNA.
Figure 2.4: Structure of the iron *E. coli* deformylase bound to formate. (A) Ribbon diagram of the secondary elements, which are color-coded with α-helices as blue, β-strands as red, and random coils as green. The atoms of the metal center are shown as ball-and-stick models and are colored by elements, with carbon as gray, nitrogen as blue, oxygen as red, sulfur as yellow, and the iron as green. (B) Electrostatic potential energy surface diagram of the protein in a similar orientation.
Figure 2.5: Ball-and-stick stereodiagrams of the metal sites of (A) iron deformylase bound to formate, and (B) zinc thermolysin bound to a phosphoramidates (PDB ID: 6TMN). Hydrogen-bonds are indicated by dashed lines. The atoms are colored by element, with carbon as gray, nitrogen as blue, oxygen as red, sulfur as yellow and the metal ion as green.
Figure 2.6: The assignment of formate. (A) $F_O - F_C$ omit density map at $6\sigma$ for active site of the cobalt deformylase grown in 3.6 M sodium formate. (B) $2F_O - F_C$ (blue, $3\sigma$) and $F_O - F_C$ (violet, $6\sigma$) density maps following incorporation of two water molecules. (C) $2F_O - F_C$ map at $3\sigma$ following incorporation of the formate molecule.
Figure 2.7: Ball-and-stick diagrams of the metal sites of formate-bound deformylases (A) iron form, (B) cobalt form, (C) zinc form, (D) superimposed iron (green), cobalt (pink), and zinc (purple) forms. The atoms are colored by element, with carbon as gray, nitrogen as blue, oxygen as red, sulfur as yellow, iron as green, cobalt as pink, and zinc as purple.
Figure 2.8: Structure of the zinc *E. coli* deformylase-PCLNA complex. (A) Ribbon diagram of the secondary structural elements that are color-coded as in Figure 2.4. The atoms of the metal center are shown as ball-and-stick models and are colored by their elements, with carbon as gray, nitrogen as blue, oxygen as red, sulfur as yellow, phosphorus as magenta, and the zinc as green. (B) Electrostatic potential energy surface diagram of the protein and stick diagram of the PCLNA inhibitor in a similar orientation.
Figure 2.9: Stereoview of the F_0-F_C electron density map of the active site of the zinc deformylase-PCLNA complex with the inhibitor omitted (blue, 5σ; red, 15σ). The carbon atoms of protein are colored in green, while those for the inhibitor are colored in purple. The remaining atoms are colored by element, nitrogen as cyan, oxygen as red, phosphorus as magenta, sulfur as yellow, and the metal ion as green.
Figure 2.10: (A) Protein-inhibitor hydrogen-bonding interactions within the zinc deformylase-PCLNA complex. (B) Interactions between inhibitor and protein. The residues located within 4.0 Å of the inhibitor are listed.
Figure 2.11: (A) Stereoview of the active site of the zinc deformylase-PCLNA complex. Hydrogen-bonds are indicated by dashed lines. The atoms are colored by element as in Figure 2.8. (B) Superimposed stereoview diagram of the structures and interactions of phosphonate inhibitors bound to thermolysin (brown) and deformylase (blue). The inhibitors of thermolysin and deformylase are colored as light yellow and dark blue, respectively. The thermolysin coordinates were obtained from the Protein Data Bank (PDB ID: 6TMN). (C) Hypothetical model for the transition state involved in the deacylation of acetylated peptides to illustrate the steric interaction of a methyl group (black) with the backbone carbonyl of Glu45.
Figure 2.11
Figure 2.12: Superposition of deformylase structures. (A) Overlap of native and formate-bound structures, *E. coli* Fe form bound to formate as green, *S. aureus* Zn form (PDB ID: 1LQW) as blue, and *P. falciparum* Co form (PDB ID: 1JYM) as pink. (B) Overlap of inhibitor-bound structures, *E. coli* Co form bound to PCLNA as pink, *E. coli* Fe form bound to PEG (PDB ID: 1BSZ) as green, *E. coli* Ni form bound to Met-Ala-Ser (PDB ID: 1BS6) as brown, and *B. subtilis* Ni form bound to Actinonin (PDB ID: 1LQY) as blue.
Figure 2.13: Proposed mechanism of deformylation by *E. coli* peptide deformylase.
<table>
<thead>
<tr>
<th>Data set</th>
<th>Fe-formate</th>
<th>Co-formate</th>
<th>Zn-formate</th>
<th>Co-PCLNA</th>
<th>Zn-PCLNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6(_2)2</td>
<td>P6(_2)2</td>
<td>P6(_2)2</td>
<td>P6(_5)2</td>
<td>P6(_2)2</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>a = b = 54.82</td>
<td>a = b = 54.42</td>
<td>a = b = 54.78</td>
<td>a = b = 98.38</td>
<td>a = b = 100.12</td>
</tr>
<tr>
<td></td>
<td>c = 224.579</td>
<td>c = 224.43</td>
<td>c = 227.27</td>
<td>c = 109.37</td>
<td>c = 111.34</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.2830</td>
<td>1.0000</td>
<td>1.2830</td>
<td>1.1355</td>
<td>1.1355</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0-1.85</td>
<td>20.0-1.30</td>
<td>20.0-1.85</td>
<td>20.0-2.00</td>
<td>20.0-2.30</td>
</tr>
<tr>
<td>Observations/Unique</td>
<td>50192/18000</td>
<td>408686/48707</td>
<td>67797/18498</td>
<td>103435/20608</td>
<td>51025/12933</td>
</tr>
<tr>
<td>Completeness (%)a</td>
<td>97.3(93.9)</td>
<td>98.2(86.9)</td>
<td>99.0(97.3)</td>
<td>90.8(93.8)</td>
<td>85.1(86.1)</td>
</tr>
<tr>
<td>Rsym (%)b</td>
<td>8.0(25.8)</td>
<td>7.1(23.9)</td>
<td>8.7(26.2)</td>
<td>7.4(31.3)</td>
<td>7.9(28.3)</td>
</tr>
<tr>
<td>No. of protein (ligand) atoms</td>
<td>1329</td>
<td>1329</td>
<td>1320</td>
<td>1359</td>
<td>1359</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>158</td>
<td>196</td>
<td>103</td>
<td>100</td>
<td>76</td>
</tr>
<tr>
<td>R(<em>{\text{crystal}}) (R(</em>{\text{free}})) (%)</td>
<td>19.9 (22.8)</td>
<td>21.9 (22.4)</td>
<td>28.0 (31.7)</td>
<td>19.9 (23.0)</td>
<td>20.1 (23.3)</td>
</tr>
<tr>
<td>B-factor of protein (metal) (Å(^2))</td>
<td>19.1 (13.3)</td>
<td>21.5 (16.7)</td>
<td>30.4 (31.2)</td>
<td>18.0 (13.5)</td>
<td>24.5 (18.0)</td>
</tr>
<tr>
<td>B-factor of product/inhibitor (Å(^2))</td>
<td>18.7</td>
<td>24.8</td>
<td>26.9</td>
<td>11.4</td>
<td>19.8</td>
</tr>
<tr>
<td>RMSD of bond distance (Å)</td>
<td>0.007</td>
<td>0.008</td>
<td>0.009</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>RMSD of bond angles (°)</td>
<td>1.200</td>
<td>1.200</td>
<td>1.200</td>
<td>1.876</td>
<td>1.680</td>
</tr>
</tbody>
</table>

a The numbers in parentheses are for the highest resolution shell.

b Rsym (I) = Σ\(_h\)Σ\(_i\) |I - I| / Σ\(_h\)Σ\(_i\) I, where I is the mean intensity of the \(_i\) observations of reflection \(_h\).

c R\(_{\text{crystal}}\) = 100 x Σ\(_i\) |F\(_{\text{obs}}\) - F\(_{\text{calc}}\)| / Σ\(_i\) |F\(_{\text{obs}}\)|, where F\(_{\text{obs}}\) and F\(_{\text{calc}}\) are the observed and calculated structure factors, respectively.

d R\(_{\text{free}}\) is the same as R\(_{\text{crystal}}\) was calculated using 8% (for native, Co-, and Zn-PCLNA) or 10% (for Fe-, Co-, and Zn-formate) of the data excluded from refinement.

Table 2.1: Data processing and refinement statistics.
CHAPTER 3

Crystal structures of the *Bradyrhizobium japonicum* FixL heme domain and ligand-bound complexes

3.1 Introduction

FixL proteins are biological oxygen sensors that regulate nitrogen fixation gene expression in *Rhizobia* (90-92). FixL proteins are comprised of two distinct domains. Located near the N-terminus is a heme-based oxygen-sensing domain belonging to the PAS superfamily (93). At the C-terminus is a histidine kinase domain that transduces its signal through a conserved phosphoryl transfer mechanism (91, 94). In the absence of dioxygen, the histidine kinase is active (“on”) and undergoes autophosphorylation with a \( \gamma \)-phosphate from ATP, followed by transfer of the phosphoryl group to its associated two-component response regulator, FixJ, which triggers a cascade of gene expression required for nitrogen fixation (95-97). Conversely, binding of dioxygen to the heme domain of FixL probably leads to a conformational change that turns “off” the histidine kinase activity. Hence, one of the fundamental questions for FixL proteins is how the binding of target ligand regulates the allostery both within the heme domain, and between the heme domain and its associated histidine

On the other hand, as is typical of heme proteins, besides the physiological ligand, dioxygen, FixL proteins can also bind to other classic heme ligands, such as CO, NO,
CN⁻, and imidazole (98, 99). It has shown that both the autophosphorylation and FixJ phosphoryl-transfer activities towards these heme ligands are dramatically different from each other. Obviously, as a sensor, FixL proteins must be able to discriminate its target ligands from those of similar size and shape, in order to induce their ligand specific functional activities. Therefore, one critical issue in the study of FixL proteins is the details for how the ligand recognition and discrimination are achieved.

To address these issues, we have chosen to work with a N-terminal truncated heme-containing domain of the FixL protein from *Bradyrhizobium japonicum* (BjFixLH). Structures of BjFixLH, in both ferric and ferrous forms (met- and deoxy-BjFixLH), as well as a series of ligand-bound complexes (oxy-, cyanomet-, imidazole-, Melm-, CO-, NO-BjFixLH), have been determined by x-ray crystallography (100, 101). These structures reveal a novel heme-binding fold, which also resembles a PAS fold conserved in many signal transduction systems. Moreover, by comparing these structures, a mechanism for ligand discrimination and allostery within the heme domain of BjFixL protein have been developed.

3.2 Materials and methods

3.2.1 Crystallization, sample preparation and data collection

BjFixLH is a monomeric 14.9 kDa protein and contains about 130 amino acids. The purified BjFixLH protein was provided by the group of Dr. Marie-Alda Gilles-Gonzalez (90, 92). Crystals of met-BjFixLH were grown at 4°C in 4.5 M NaCl / 5% 2-methyl-2,4-pentanediol (vol/vol) / 0.1 M Hepes buffer (pH 7.5) by the vapor diffusion method and achieved their final size within 1 month (Figure 3.1). A platinum derivative
was obtained by soaking native crystals with a mixture of 5 mM di-µ-iodobis(ethylenediamine)diplatinum (II) nitrate and 5 mM potassium tetracyanoplatisnate for 4 days. Deoxy-BjFixLH crystals were obtained by soaking met-BjFixLH crystals with degassed synthetic mother liquor containing saturated sodium dithionite.

The ligand-bound crystals were prepared by soaking the met-BjFixLH crystals under various conditions. Oxy-BjFixLH crystals were generated by soaking met-BjFixLH crystals with solutions of synthetic mother liquor containing saturated O₂ and 100 mM ascorbic acid. Cyanomet-BjFixLH crystals were obtained by either directly growing the crystal with 10 mM potassium cyanide or by treating met-BjFixLH crystals with 100 mM potassium cyanide for 30 minutes. Imidazole- and MeIm-BjFixLH crystals were prepared by soaking met-BjFixLH crystals with solutions of synthetic mother liquor containing 20 mM imidazole or 200 mM 1-methylimidazole for 90 minutes. NO-BjFixLH crystals were obtained by soaking met-BjFixLH crystals with glycerol followed by addition of diethylamine NONOate (Cayman Chemical). The NO-BjFixLH crystals were then sealed under positive NO pressure for one day. All the above different crystal forms were transferred stepwise to solutions of synthetic mother liquor containing increasing amounts of glycerol up to a maximal concentration of 40% (vol/vol) prior to being flash-cooled in liquid nitrogen.

CO-BjFixLH crystals were generated by using a MRC Cryo-Xe-Siter at the beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL). The met-BjFixLH crystals were transferred stepwise to anaerobic synthetic mother liquor containing increasing concentrations of glycerol (up to 30% (vol/vol)) and then soaked in a solution containing saturated sodium dithionite for several minutes. The reduced crystals were
then mounted on a cryo-loop and pressurized in the Cryo-Xe-Siter under a CO pressure of 300 psi for 70 minutes. The CO-BjFixLH crystals were flash-cooled under pressure in liquid carbon tetrafluoride. They were then flash-cooled in liquid nitrogen for data collection. Upon treatment, the color of both ferrous and ligand-bound crystals changed from orange-red (met-BjFixLH) to cherry-red within couple of minutes.

All of the diffraction data were collected at 100 K at one of these synchrotron facilities (Beamline X4A at the National Synchrotron Light Source in Brookhaven National Laboratory (BNL); Beamline BIOCARS 14D at the Advance Photon Source (APS) at Argonne National Laboratory; Beamline 9-1 at SSRL.) on either an RAXIS 100 or ADSC-Q4 CCD detector (Table 3.1 and 3.2). Data processing and reduction were performed using DENZNO and SCALEPACK (102). The crystals belong to the space group R32 with unit cell dimensions \( a = b = 128.8 \text{ Å}, c = 58.9 \text{ Å} \), and contain one BjFixLH per asymmetric unit. Multiwavelength anomalous diffraction (MAD) data for met-BjFixLH were collected at four different wavelengths, corresponding to the inflection point of the Fe absorption edge (1.7396 Å) (Nat1), the maximum of the peak above the edge (1.7379 Å) (Nat2), and two remote points at higher energy (1.6984 Å and 1.4938 Å) (Nat3 and Nat4).

3.2.2 Structure determination and refinement

The structure of met-BjFixLH was solved by the combination of the MAD and multiple isomorphous replacement and anomalous scattering (MIRAS) methods. The phases were initially determined by the program PHASES (103) by using the data collected near the iron edge to take advantage of the anomalous scattering from the
mononuclear iron site (Table 3.1). A platinum derivative data was also collected to provide conventional isomorphous and anomalous phases (Table 3.1). Both the MAD and derivative data were required to obtain interpretable maps. The heme iron site was located from the Nat2 anomalous Patterson map and confirmed from the Nat3-Nat1 difference Patterson map (Figure 3.2). MAD phasing was carried out by treating the data as multiple isomorphous replacement and anomalous datasets: native anomalous scattering (Nat1-ano), derivative anomalous (Nat2-ano, Nat3-ano, and Nat4-ano), and derivative isomorphous scattering (Nat1-Nat3, Nat1-Nat4). For the platinum derivative, three Pt sites with low occupancy were located by difference Fourier methods using the initial MAD phase and then were included in the final phase calculation and refinement. The phases were further improved by solvent flattening with the PHASES program. At this stage, the electron density was clear enough to locate the heme and trace most of the backbone. The model building and structure refinement were carried out by the programs O (68, 104) and X-PLOR (105), respectively. Eight percent of the data were set aside for calculation of the free-R factor (70). The first twelve residues at the N terminus of the met-BjFixLH were not observed and are presumably disordered. A flat bulk solvent correction and an overall anisotropic B-factor scaling were applied to the data. Solvent molecules were found from the 3σ Fo-Fc map. Pro 195 was found to be a cis-proline. The met-BjFixLH structure was refined to 2.4 Å resolution.

The structures of deoxy-BjFixLH and each of the ligand-bound complexes were determined using the met-BjFixLH structure (PDB entry: 1DRM) as the starting model. Model building and refinement were performed using the programs O (104) and X-PLOR 3.851 (70), respectively (Table 3.2). The ligands, solvent molecules, and the residues
involved in the conformational change were modeled based on the resulting 2F_O-F_C and F_O-F_C electron density maps. The quality of the model for each structure was evaluated using PROCHECK (71). All stereochemical parameters for each structure were within acceptable ranges with majority being better than acceptable. Figures were prepared with the programs Molscript (74), Raster3D (75) and XtalView (106). The root mean square deviations (RMSD) after least square fitting of the overall, main-chain, and heme atoms of deoxy-BjFixLH and each of the six ligand-bound structures (CO-, NO-, cyanomet-, imidazole-, MeIm-, and oxy-BjFixLH) to met-BjFixLH were determined using the program CNS (69). The least square alignment using the coordinates of four PAS domains, PYP, HERG, LOV2 and deoxy-BjFixLH (PDB ID codes 2PHY, 1BYW, 1G28, and 1DRM, respectively) were performed by program TOP (107)

3.3 Results and discussions

3.3.1 The overall heme-domain fold

All forms of BjFixLH (unliganded and liganded) adopt the same basic fold (100, 101). Unlike the more commonly studied classes of heme proteins (such as myoglobin, cytochrome c, cytochrome P450 CAM) that are mainly alpha-helical, the dominant structural feature of the BjFixL heme domain is a five-stranded antiparallel sheet (Figure 3.3). The secondary structure is assigned by PROCHECK (71) and labeled in alphabetical order: β strands: A_β, A155-A161; B_β, A164-A168; G_β, A219-A225; H_β, A231-A243; I_β, A246-A255; helices: C_α, A170-A176; D_α, A180-A184; E_α, A188-A191; F_α, A196-A210; J_α, A257-A270. Each loop would be defined by the secondary structures that flank it (e.g., FG loop). Using this nomenclature, the overall fold of the BjFixLH can be
described as a left-handed glove that encloses a heme cofactor. The fingers are formed from \( \beta \) strands, \( H_\beta, I_\beta, A_\beta, \) and \( B_\beta \), the palm from \( \alpha \) helices, \( C_\alpha, D_\alpha, E_\alpha, \) and \( F_\alpha \), and the thumb from \( \beta \) strands, \( C_\beta \), and \( H_\beta \). The only distortion from this closed hand is the presence of a C-terminal helix, \( J_\alpha \), which protrudes from one of the \( \beta \) strands of the fingers and leads into the kinase domain. This glutamine-rich helix, called the Q-linker, is not part of the domain proper, and instead serves as the linker between the heme and kinase domains. The heme is cusped within the left side of the glove. Recently, crystal structures of FixL heme domain from \textit{Rhizobium meliloti} (RmFixLH) have also been solved and reveal a virtually identical overall fold as of BjFixLH (108).

### 3.3.2 The PAS fold

As mentioned earlier, another interest in the heme domain of FixL proteins stems from the fact that sequence alignments suggest that it shares homology with several PAS domain proteins (93). PAS domains are named after the gene products of the initial three members identified: \textit{Drosophila} period (\textit{Per}), vertebrate aryl hydrocarbon receptor nuclear translocator (\textit{Arnt}), and \textit{Drosophila} single-minded (\textit{Sim}) (109). PAS domains occur in many sensory proteins among Bacteria, Archaea, and Eukarya (93, 110, 111). They contain a small number of conserved amino acids within a sequence of about 90 residues (Figure 3.4). It has been known that many PAS-domain proteins detect their signals by way of an associated cofactor (90, 112, 113). Their signal transduction is likely mediated and carried out by protein-protein interactions.

To date, besides BjFixL heme domain, there are four other crystal structures from the PAS superfamily that have been determined: the bacterial blue-light photoactive
yellow protein (PYP) with a covalently attached 4-hydroxycinnamoyl chromophore ((110), the N-terminal regulatory domain of the human ether-a-go-go-related gene potassium ion channel (HERG) (114), the LOV2 domain of the phototropin module of *Adiantum* phy3 containing flavin mononucleotide (FMN) (115), and the heme domain of *Rhizobium meliloti* FixL (108). Despite the distinct functions and the overall low sequence similarity, surprisingly, these proteins reveal significant structural homology within their PAS domains (Figure 3.5A). The RMSD of 70 matched residues from these proteins to met-BjFixLH are 1.39 Å to PYP, 1.49 Å to HERG, 1.06 Å to PYP, 0.62 Å to RmFixLH, whereas the sequence identities among them are no more than 20%. Generally, the PAS fold consists of three parts: (1) a PAS core, made up of the two N-terminal β-sheets and three helices, (2) a helical connector, composed of the long α helix that diagonally crosses the β-sheets, and (3) a β-scaffold, composed of the three long β-strands that comprise the second half of the central β-sheet (Figure 3.5B) (110).

The major structural difference among these PAS proteins, however, occurs at the region around the central F helix. Like the palm of a hand, this helix, together the EF loop preceding it, shifts to accommodate various cofactors within the domain. In BjFixLH and RmFixLH, the F helix (palm) adopts a position that sandwiches the heme between itself and the GH strands (the thumb). In the phy3 LOV2 domain, this helix shifts away from the C, D, E helices, to form a channel for the FMN cofactor. The chromophore in PYP lies on opposite sides of the F helix and interacts with the EF loop. These results point to the importance of the central helix F and the loops that flank it as being the critical regulatory region for the PAS domain family.
3.3.3 The heme pocket

Both deoxy- and met-BjFixLH contain a pentacoordinate heme iron, with the imidazole side chain of His 200 providing its axial ligand and no other ligands coordinating the heme. As expected for pentacoordinate hemes, the porphyrin ring appears significantly puckered. The heme-propionate 6 forms hydrogen bonds with the backbone NH's of residues 214-216, whereas the heme-propionate 7 forms salt bridges with the side chains of His 214 and Arg 220 (Figure 3.6). The heme pocket is primarily hydrophobic. Residues Val 188, Met 192, Tyr 203, and Ile 204 help to orient the His 200 on the proximal side of the heme. Two additional residues on the same side, Ile 159 and Leu 191, are more peripheral. The side chains of Ile 157, Phe 176, Gly 224, Phe 249, and Tyr 207 lie in the same plane as the porphyrin ring. Ligands bind to the heme at the open axial site trans to the proximal histidine in a distal pocket defined by Ile 215, Leu 236, Ile 238, Val 222, Met 234, and Ile 216. Even though the ligand-binding site is hydrophobic, it is fairly accessible by means of an entryway marked by three water residues that interact with Arg 220. The hydrophobic interior near the ligand-binding site is bounded by the side chains of Ile 215, Leu 236, and Ile 238 (Figure 3.7A).

Regarding the overall secondary structures and the regions surrounding the heme, considerable differences have been revealed in comparison studies between the BjFixLH structure and that of myoglobins and hemoglobins. Therefore, one would predict little or no similarity in their binding sites. However, some surprising similarities have been shown in the alignment of the central heme and the bound histidine of BjFixLH and *Glycera dibranchiata* hemoglobin, which also has a distal (E7) leucine instead of the more common histidine (Figure 3.8) (116). Two of the hydrophobic groups on a strand
within the distal pocket of BjFixLH, Leu 236, and Ile 238 are positioned analogously to Leu 58 (E7) and Val 62 (E11) of *G. dibranchiata* hemoglobin, respectively. The other residue in BjFixLH that is close enough to interact with bound ligands, Ile 215, is part of a critical loop involved in the regulatory signal. The presence of this additional residue results in a slightly smaller hydrophobic distal pocket for BjFixLH. The close similarity of the relative orientations of the histidine ligand and the two distal side chains, in these two clearly evolutionarily distant oxygen binding proteins, raises the question whether there are also electronic effects associated with oxygen binding that nature is able to exploit. Consistent with these observations, above features have also be found in the ligand-binding pocket of the nitric-oxide (NO) transport protein from a blood-sucking insect (Figure. 3.8) (117). When their coordinated histidines are oriented similarly, Leu 236 and Ile 238 in BjFixLH are positioned similarly to Leu 123 and Leu 133, respectively, in the NO transport protein. However, the side chain residue of Leu 123 in the NO transport protein is farther away from the metal than Leu 236 in BjFixLH. This difference may reflect the fact that there are different functional ligands for the two proteins.

3.3.4 Probing the role of the oxidation state of the heme iron on the protein conformation

Historically, it has been thought that the oxidation state of the heme iron in FixL has no effect on the structure of the heme domain because the autophosphorylation activities of both ferric and ferrous forms of FixL were found to be the same (118). However, recently studies have shown that the turnover of RmFixJ to phospo-RmFixJ
by RmFixL is dependent on the oxidation state of the RmFixL heme, with the ferrous form being 100 times more active (99). While this suggests the possibility of redox-dependent differences in the structure of RmFixL, the structures only exhibit minor difference between ferrous deoxy-RmFixLH and ferric met-RmFixLH. One possibility, however, is that since no ligand-bound forms of RmFixLH could be prepared, crystal packing forces could be stabilizing the conformation observed in both oxidation states. As conformational changes had been observed in our BjFixLH structures, both the structures of deoxy- and met-BjFixLH were determined and compared to examine this issue.

Met-BjFixLH crystals were obtained from the initial crystallization trials. Deoxy-BjFixLH crystals were prepared by soaking met-BjFixLH crystals with sodium dithionite. A rapid color change was observed. The overall structures of deoxy- and met-BjFixLH are virtually identical (Figure 3.9). Based on main-chain atoms, deoxy-BjFixLH has the smallest RMSD to met-BjFixLH of all of the BjFixLH structures we have determined (Figure 3.10). The heme-binding pocket for deoxy-BjFixLH also overlaps with that of the met-BjFixLH (Figure 3.7A and B). Around the heme pocket, the only subtle difference lies in the interaction between the heme-propionate 6 and Arg 206. In met-BjFixLH, the heme-propionate 6 and Arg 206 are hydrogen-bonded via a water molecule. In deoxy-BjFixLH, this water molecule is missing and this long-range interaction is lost. The similarity of these structures suggests that if there are oxidation state differences in the conformation of the heme domain, either the kinase domain and/or complexation with FixJ, is required to induce these differences.
Although our structures have shown no differences between the ferric and ferrous states of the heme domain, one observation that may be consistent with the presence of certain subtle differences in their properties, has been the apparent difficulty in preparing ligand-bound forms of ferrous BjFixLH. This contrasts the relative ease of preparing ligand bound forms of ferric BjFixLH. While the origin of these differences is unclear, we note that this could partially due to the fact that ferric iron is smaller than ferrous iron and thus would lead to flattening of the heme, which would facilitate the binding of ligands to the open axial site. We note that if such an oxidation-state dependence existed, it would promote the binding of O₂, which can partially oxidize the heme iron, relative to ligands, such as CO and NO, that do not. Clearly, additional studies are required to answer the question whether oxidation state is indeed vital in facilitating ligand discrimination of O₂.

### 3.3.5 The conformation change in oxy-BjFixLH

One of the fundamental issues in regard to FixL is the signal transduction mechanism. To elucidate the molecular details of this mechanism, a series of ligand-bound complex structures of BjFixLH were determined. The overall fold of these ligand-bound forms of BjFixLH was found to be similar to that of the deoxy- and met-BjFixLH. For some of the ligand-bound forms (oxy-, cyanomet-, imidazole-, and MeIm-BjFixLH), however, a conformational change was observed in one region of the proteins (Figure 3.10A). Since dioxygen is the physiological ligand of BjFixL, the structure of oxy-BjFixLH is perhaps the most important form of these, and thus we will begin by describing its structure.
Oxy-BjFixLH crystals were obtained by treating met-BjFixLH crystals with ascorbic acid (to reduce the heme iron) in the presence of saturating dioxygen. The oxy-BjFixLH structure is shown in Figure 3.11A. The 2Fo-Fc and Fo-Fc electron density maps revealed density in the heme pocket, which is consistent with a dioxygen ligand that binds the heme iron in a bent mode (\( \angle \text{Fe-O-O} \approx 124^\circ \)) (Figure 3.11B). The plane formed by the heme iron and the bent dioxygen ligand appears to be aligned with the plane of the proximal histidine ring.

Comparison of the deoxy- and oxy-BjFixLH reveals distinct changes in the heme plane and a loop region between helix F and strand G, termed the FG loop (Thr 209 and Arg 220) (Figure 3.12A). Upon binding of dioxygen, the porphyrin ring of the heme becomes more planar, and two heme-propionate side chains shift to positions that are farther from the heme iron (Figure 3.12B). The first carbon of heme-propionate 7 (CAA), and the pyrrole carbon to which it is attached (C2A), for example, move by 1.10 and 0.83 Å, respectively. As the expected error in the atomic positions in these structures is approximately 0.3 Å at the current resolution (2.3 Å), this shift is fairly dramatic.

Meanwhile, the interactions between heme-propionate arms and the protein are altered (Figure 3.12, 3.7B and E). The most significant change occurs in the conformation of Arg 220, a residue that is conserved in all FixL proteins. While Arg 220 in met-BjFixLH forms a hydrogen bond with heme-propionate 7 O2A, the guanidium side chain of Arg 220 in oxy-BjFixLH rotates into the heme pocket and forms a hydrogen bond with the bound dioxygen ligand. This change is fairly remarkable, requiring the C\(_{\alpha}\)-C\(_{\beta}\) bond of Arg 220 to rotate by \( \sim 170^\circ \), and it suggests that Arg 220 may play an important role in FixL’s O\(_2\) sensing mechanism. Another conformational change in the
FG loop occurs to His 214. In deoxy-BjFixLH, the Nδ1 atom of His 214 forms a salt bridge primarily to the propionate O1A, whereas in oxy-BjFixLH, this same nitrogen shifts about 1.6 Å away from the heme, and interacts solely with the propionate O2A. The contribution of heme-propionate 6 to the conformational change is much subtler. Hydrogen bonding of the O1D propionate oxygen shifts from the main-chain NH groups of residues 214 – 216 to the main-chain NH groups of residue 214 and 215. The O2D propionate oxygen, which hydrogen bonds to the main-chain NH of His 214 and a water in deoxy-BjFixLH, shifts its interaction to the Nδ1 of Arg 206, as well as a water, in oxy-BjFixLH.

Speaking of Arg 206, in deoxy-BjFixLH, this residue undergoes hydrogen bonding with the main-chain carbonyl of Asp 212, the side-chain carboxylate of Asp 212, and the side-chain Nδ2 of His 214, respectively. In oxy-BjFixLH, Arg 206 maintains its hydrogen bond to the Asp 212 main-chain carbonyl, but it can no longer form a hydrogen bond with the side chain of Asp 212 and His 214 because of its interaction with the heme propionate 6. Because of the loss of the salt bridge to Arg 206, Asp 212 rotates around its C-C bond so as to form a weak interaction with the backbone carbonyl of Thr 210.

Another important aspect of the conformational change can be found in the distal pocket of the oxy-BjFixLH structure (Figure 3.7B and E). Concomitant with the shift of the FG loop is a reorientation of the main-chain atoms of Ile 218. While the carbonyl of Ile 218 in met-BjFixLH points towards the solvent, as a result of the conformational change, the carbonyl of Ile 218 in oxy-BjFixLH rotates into the heme pocket forming a hydrogen-bonding interaction with a water molecule. In return, this water forms a hydrogen bond with Arg 220 and may help to stabilize the arginine in a position suitable
for hydrogen-bonding formation with the bound dioxygen ligand. Furthermore, the hydrophobic side chains of Ile 215, Leu 236, and Ile 238 adjust in oxy-BjFixLH in order to accommodate the bound dioxygen.

### 3.3.6 Comparison of the Hydrogen bonding in BjFixLH and other O$_2$-binding heme proteins

One of the fundamental issues regarding FixL proteins has been whether they utilize a hydrogen-bond donor to stabilize the binding of dioxygen to the heme. Historical precedent for hydrogen-bonding interactions in an oxygen-binding heme-protein has been demonstrated in classical myoglobins and hemoglobins that utilize a distal histidine (residue E7) to anchor the bound dioxygen. Even in *Aplysia* (119) and elephant myoglobins (120), both of which lack the E7 histidine, a stabilizing interaction is provided by an E10 arginine or B10 glutamine, respectively. The oxy-BjFixLH structure has revealed that Arg 220 serves as the distal ligand that can stabilize the bound dioxygen through the hydrogen bonding. Despite the different conformation around heme pockets and the identities of the distal ligands, superposition of the heme cofactors in oxy-BjFixLH and in the 1.0 Å structure of sperm whale oxy-myoglobin (PDB entry: 1A6M) reveals similar orientations of the heme, proximal histidine, and distal ligand relative to the bound dioxygen ligand (Figure 3.13). The aromatic ring of the proximal histidine and the plane of the bent (Fe-O-O) dioxygen ligand are approximately aligned, while the hydrogen bond from the distal ligand appears to originate from a similar orientation. The conservation of these features in distinctively different heme proteins supports the notion
that the orientations of the dioxygen ligand, the proximal histidine, and the distal ligand all contribute to dioxygen binding affinity (121).

### 3.3.7 Probing the role of spin-state and ligand sterics on the conformational change by the structures of NO- and CO-BjFixLH

The feature that drives the heme-mediated conformational change in FixL proteins has been an area of great speculation. One of the popular mechanisms when we started this project was the proposal that the change in spin-state of the heme iron was the driving force for the ligand-induced conformational change. This mechanism was based on the observation that the binding of the strong-field ligands for both ferric and ferrous oxidation states inhibits the kinase activity of FixL (118). The kinase is active when the heme iron is high-spin, as in unliganded ferrous or ferric FixL, whereas the binding of strong-field ligands (e.g. dioxygen, cyanide) changes the heme iron to low-spin and inactivates the kinase. It has been demonstrated that in hemoglobins, the change of the heme-iron from high-spin to low-spin upon dioxygen binding plays an important role in the allosteric mechanism (10, 122). Due to its smaller size, the low-spin iron can move into the plane of the porphyrin ring, thus causing a shift of its attached proximal histidine and its associated F-helix.

Consistent with the spin-state model, there has been speculation that nitric oxide, also a strong-field ligand, could serve as a secondary ligand even though dioxygen is clearly the physiological ligand for FixL proteins (98). In support of this hypothesis, nitric oxide has been noted to be the potential product of denitrification by *Rhizobia* (123). In addition, sensing of nitric oxide by heme proteins has been demonstrated for the
soluble guanylyl cyclase of vertebrates (124). If nitric oxide is indeed a physiologically relevant ligand, it should be able to induce the same conformational changes as dioxygen.

More recently, the steric of the ligand has also been taken into the consideration as the cause of the conformational change based on data suggesting a tight heme distal pocket and the FixL's fast on-rates for imidazole binding (125). This model has also been supported by mutagenesis experiments, but these data appear inconclusive. Mutation of the distal residues Ile 209 and Ile 210 in the FG loop (corresponding to Ile 215 and Ile 216 in BjFixL) to Ala, His and Trp have been shown to lead to an increase in the kinase autophosphorylation activity for both deoxy- and oxy-RmFixLT (126). This result has been used to suggest that these residues play an important role in FixL's O₂ sensing mechanism. The steric interaction between these residues and the bound-dioxygen ligand is proposed to be the driving force of the conformational change. While this argument makes sense for the Ala mutations, the data appear inconsistent for the His and Trp mutants. Both His and Trp are larger than Ile and thus should lead to down regulation if the steric hindrance model is correct. To test both the spin-state and ligand steric models, we have determined the structures of NO- and CO-BjFixLH.

NO-BjFixLH crystals were prepared by treating met-BjFixLH crystals with diethylamine NONOate in a sealed capillary. The 2Fo-Fc and Fo-Fc electron density maps reveal density in the distal pocket consistent with the presence of a bent nitrosyl ligand bound to the heme iron (\( \angle \text{Fe-N-O} \approx 146^\circ \)) (Figure 3.7D). However, more importantly, unlike oxy-BjFixLH, there is no major conformational change of the FG loop observed in NO-BjFixLH. Comparison of the C\(_{\alpha}\) atoms within the critical FG loops of met-BjFixLH and NO-BjFixLH yields an average positional difference of only 0.1 Å.
The two heme-propionate arms remain in approximately the same positions as well. In heme pocket, the side chain of Ile 238 simply adjusts to accommodate the bound NO ligand. These results appear to suggest that nitric oxide is not a secondary ligand of BjFixLH. Moreover, it raises doubts for the roles of spin-state and ligand sterics in the conformation change.

A detailed examination of the NO-BjFixLH structure reveals that although there is no shift of the FG loop, the binding of nitric oxide to the heme still affects the heme ruffling. The nonplanar heme-distortion in NO-BjFixLH appears to be halfway between that of met- and oxy-BjFixLH. Addition of NO results in a shift of 0.44 Å for the first carbon of heme-propionate 7 (CAA) and a shift of 0.28 Å for the pyrrole carbon to which heme-propionate 7 is attached (C2A). Moreover, there is another subtle different around the heme pocket. In the unliganded met-BjFixLH structure, the guanidium side chain of Arg 206 interacts with the main-chain carbonyl and side-chain carboxylate of Asp 212. In the oxy-BjFixLH structure, which exhibits the typical conformational change, Arg 206 appears to shift slightly to a position that interacts with the carbonyl of Asp 212 and the carboxylate of heme propionate 6. On the other hand, the position of Arg 206 in NO-BjFixLH differs from either of these structures, and instead, the side-chain of Arg 206 is found to form a salt-bridge with the heme propionate 7. There is no other apparent change in the polypeptide. These observations lead us to suggest that the NO-BjFixLH structure may model an intermediate in the heme-driven conformational change. Further studies are required to address this issue.

The unexpected results from NO-BjFixLH structure, however, could be due to the unusual property of the NO-bound hemes since they tend to be distorted (127, 128).
Heme bound to CO, on the other hand, should be low spin and yet adopt more typical geometries based on previously determined structures of globins and porphyrin model compounds (129, 130). Hence, as a further test, the structure of ferrous CO-BjFixLH was determined.

It was surprisingly difficult to obtain suitable crystals of CO-BjFixLH in light of the ease of preparing crystals of oxy-BjFixLH. After numerous trials, suitable crystals were obtained via the anaerobic reduction of met-BjFixLH crystals with dithionite followed by a long-term incubation under a high pressure CO atmosphere. The structures of the overall protein and the heme-binding pocket for CO-BjFixLH are shown in Figure 3.14A and 3.7C, respectively. Importantly, $F_O-F_C$ and $2F_O-F_C$ electron density maps clearly show the presence of the CO molecule (Figure 3.14B), but there is no shift of the FG loop. The overall RMSD of the main-chain atoms between CO-BjFixLH and met-BjFixLH is only 0.23 Å, which is the smallest deviation observed for all the ligand-bound forms that have been determined so far (Figure 3.10B). Analysis of the side chains reveals only small differences in the positions of Asp 212 and Ile 238. The side chain of Asp 212 is shifted away from its salt-bridge with Arg 206 (Figure 3.8A and C), while Ile 238 is in the same position as found in oxy-, and NO-BjFixLH, which probably results from the steric interaction required to accommodate the CO ligand. Because CO is a strong-field ligand and its binding does not induce the conformational change, these data, together with the results from the structure of NO-BjFixLH, suggest that neither steric of the ligand itself nor the change of the spin-state of the heme iron is sufficient to drive the conformational change.
3.3.8 Investigating the role of electrostatic interaction in the conformational change from the structure of cyanomet-BjFixLH

Kinetic data has shown that cyanide induces the a signaling response similar to dioxygen; its affinity ($K_d \approx 10^{-6}$ M) for the heme iron is in fact greater than that of oxygen ($K_d \approx 10^{-4}$ M for BjFixLH) \(^{(98, 118)}\). In addition, cyanide has a size similar to dioxygen and could bind to ferric heme iron. Therefore, the structure of cyanide-bound BjFixLH (cyanomet-BjFixLH) would be important to determine how cyanide binding differs from CO and NO. This would help us to further understand the allosteric and ligand discrimination mechanism of BjFixLH.

The cyanomet-BjFixLH crystals were produced either by soaking met-BjFixLH crystals in high concentration cyanide or by crystallizing in the presence of cyanide to ensure complete binding. The two cyanomet-BjFixLH structures are virtually identical. The binding of cyanide leads to a shift of the FG loop and movement of Arg 220 that is similar to that observed in oxy-BjFixLH (Figure 3.14C and 3.7F). In both cyanomet-BjFixLH structures, the binding of the strong-field ligand alters the nonplanarity of the heme resulting in a shift in the positions of the two heme-propionate side chains. As a result of these shifts, Arg 220 loses its salt-bridge to heme-propionate 7 and rotates into the heme pocket to form an electrostatic interaction with the bound ligand, while Arg 206 moves to form a salt-bridge with heme-propionate 6.

This observation is significant because it provides support for the correlation of the shift of the FG loop and the rotation of Arg 220 into the heme pocket to presumably serve as a steric barrier for the conformational change. To accommodate Arg 220 in the heme pocket, Ile 215 must move away from the heme in a manner that is consistent with
the FG loop shift (Figure 3.7). In the CO- and NO-BjFixLH structures, there is no conformational change observed and Arg 220 remains in the same position as in met-BjFixLH, presumably because the electrostatic interaction between Arg 220 and these ligands is weaker. This would be consistent with the relative electrostatic stabilizations provided by the distal histidine in myoglobins (131).

These data suggest the possibility that hydrogen bonding is used by FixL to discriminate O₂ from similarly sized heme-ligands such as CO and NO. Hydrogen bonding to a bound dioxygen ligand by either a distal histidine, arginine, or glutamine is also a conserved feature for globins (10, 119, 120, 122), suggesting that hydrogen bonding is a common feature utilized by all O₂-binding heme proteins to discriminate O₂ from other ligands. It should be noted that Arg 220 must be released from its salt-bridge with heme-propionate 7 before it can rotate into the heme pocket. This feature is presumably facilitated by a flattening of the heme induced by the binding of strong-field ligands (Figure 3.10B). Thus FixL may also discriminate dioxygen from other ligands by its ligand field strength.

3.3.9 Evaluating the importance of the electronic interaction and a hydrogen bond in the heme distal pocket from the structures of imidazole- and MeIm-BjFixLH

Binding of imidazole to the FixL heme has been shown to turn off the kinase activity (99). While the results from the structures of NO-, CO- and cyanomet-BjFixLH suggest that the electrostatic interaction between the bound-ligand and a distal ligand Arg 220 is essential for the conformational change, one problem with this is that the bound-
imidazole is either neutral or positively charged, and thus could not form this hydrogen bond with Arg 220. To understand the basis of how imidazole is able to induce the conformational change, the structure of imidazole-BjFixLH was determined.

The structure of the imidazole-BjFixLH heme pocket is depicted in Figure 3.7G. The 2F_o-F_C and F_o-F_C electron density maps clearly show density consistent with a bound imidazole ligand. As observed for the dioxygen ligand in oxy-BjFixLH, the plane of the imidazole ring appears aligned with the plane of the proximal histidine. Moreover, the structure of imidazole-BjFixLH reveals changes in the FG loop that are similar to those observed in oxy-BjFixLH. Comparison of these structures reveals that the shift of the FG loop in imidazole-BjFixLH is only slightly less than that for oxy-BjFixLH (Figure 3.10). As in oxy-BjFixLH, a water molecule located within the distal pocket interacts with the carbonyl of Ile 218. On the other hand, unlike oxy-BjFixLH, in imidazole-BjFixLH, the water hydrogen-bonds the imidazole ligand directly, while the guanidium side chain of Arg 220 adopts a position outside of the heme pocket (Figure 3.7G). At first, these results seem in conflict with the critical role of the hydrogen bond between the bound-ligand and Arg 220 for inducing the conformation change. However, when the structures of imidazole- and oxy-BjFixLH were examined carefully, we noted that the imidazole ligand was positioned in an orientation similar to the arginine guanidinium group and thus could serve as its own steric barrier to stabilize the “off” state of the enzyme (Figure 3.7E and G). Therefore, this datum in fact provides support for the importance of hydrogen bonding to the bound ligand for inducing the FG loop shift.

In addition, due to its bulky size, binding of imidazole to the heme requires a reorientation of the distal residues in order to accommodate the larger ligand. Some room
is provided by a change in the rotamer of Ile 238, but the majority of the space appears to be provided by a 2.2 Å translation of Ile 215 that is associated with the FG loop shift away from the heme pocket. The favorable hydrogen-bonding interactions of imidazole with the water molecule that lies within the distal pocket may also contribute to the affinity of the BjFixLH for imidazole.

When examining the possible role of Arg 220 and exogenous imidazole in driving the observed conformational change, we noted another factor that could induce the conformational change. In each of the "off" state structures that we have determined, Arg 220 or imidazole was found to form a hydrogen bond with a water tightly bound in heme-pocket (W1 in Figure 3.7). This hydrogen bond completes a network interaction potentially linking the heme iron and the FG loop, and more importantly, the interaction is not present in any "on" state structures. To explore the significance of this hydrogen bond to the conformational change, the structure of BjFixLH bound to 1-methylimidazole was determined. 1-Methylimidazole is analogous to imidazole except that the NH hydrogen is replaced by a methyl group, and thus is incapable to form hydrogen bond with the heme-pocket water.

The structure of MeIm-BjFixLH was determined at a resolution of 2.7 Å. The ligand binds in a similar fashion to imidazole except that the orientation of the imidazole is rotated by 180° around Fe-N bond. This directs the methylated nitrogen away from the heme pocket and places the C5 carbon in the position where hydrogen bonding usually occurs (Figure 3.7H). This ring carbon cannot form a hydrogen bond with the conserved water in the heme pocket. The structure reveals that despite the loss of the hydrogen bond
to the water, the conformational change still takes place. Therefore, these data rule out the direct significance of this hydrogen bond in the conformational change.

3.3.10 A proposed mechanism for allostery and ligand discrimination within BjFixLH

Comparison of oxy-BjFixLH structure with ligand-free deoxy-BjFixLH structure reveals a heme-driven conformational change that is distinct from that of classical globins. Unlike hemoglobin, the movement of the proximal histidine and its associated α-helix does not play a major role in FixL’s allostERIC mechanism. Instead, binding of dioxygen to the FixL heme results in the flattening of the heme porphyrin and the shift of a critical loop, the FG loop, away from the heme pocket. This FG loop shift presumably induces a global conformational change in the full-length protein leading to the inhibition of the kinase activity. While cyanide, imidazole and 1-methylimidazole were found to also induce this conformational change, CO and NO did not.

Based on these structures (met-, deoxy, oxy-, CO-, NO-, cyanomet-, imidazole-, and MeIm-BjFixLH), we propose a model for ligand discrimination and allostery for BjFixL (Figure 3.15). In this model, binding of strong-field ligands to the heme initiates the process towards the conformational change by inducing a flattening of the heme plane. This helps to reduce the strength of the salt-bridge between the critical arginine (Arg 220) and heme-propionate 7 (Figure 3.15, step 1), and provides a mechanism for discriminating strong-field ligands, such as dioxygen, from weak-field ligands. Consistent with the importance of this salt-bridge, a second arginine, Arg 206 in BjFixL, is found to be a conserved positively charged residue in all PAS heme sensors. When the
heme flattens, this arginine can form a salt-bridge with the same heme-propionate 7 that interacts with Arg 220. This feature has been observed in the NO-BjFixLH structure, a possible model for the intermediate between the “on” and “off” states.

The weakening of the Arg 220/heme-propionate 7 salt bridge has two possible outcomes. If Arg 220 cannot hydrogen bond the bound ligand within the heme pocket, as for CO and NO, or the ligand itself cannot provide its own steric barrier (i.e. imidazole), there will be no conformational change. However, if Arg 220 can form a hydrogen bond to the bound ligand (O2, CN\textsuperscript{−}), Arg 220 rotates into the heme pocket, as a steric hindrance against Ile 215. This change stabilizes the "off" state of the enzyme (Figure 3.15, step 2). As the final step in this mechanism, Arg 206 shifts to form a salt-bridge to heme-propionate 6 (Figure 3.15, step 3).

3.4 Conclusions

In this work, various structures of the *Bradyrhizobium japonicum* FixL heme domain have been determined in the absence and presence of specific ligands. These structures reveal a novel heme-binding fold, and along with other four PAS domain structures, provide structural prototypes for the PAS-domain superfamily. They shed light on the versatility of the PAS motif as a switching mechanism in signal transduction systems and the interactions that occur between this class of sensory domains and the histidine kinase transmitter domains.

On the other hand, this study reveals a new mechanism for the initiation of a conformational change by a heme center, which is distinct from hemoglobin. Unlike hemoglobin, where movement of the axial histidine ligand induces the well-known
allosteric changes, the axial ligand is rigidly held in FixL. Instead, it is the flattening of
the porphyrin itself that triggers the conformation change.

Despite these differences, the elements involved in driving the allostery and
ligand discrimination appear to be the same: spin-state, electrostatics, and sterics. Based
on our current model, the ligand-field strength of dioxygen is used to alter the spin-state
of the heme iron that in turn initiates a series of events that results in the release of Arg
220 from its salt-bridge to heme-propionate 7. This arginine can then be used to
discriminate O₂ from CO and NO by differentiating the ability of these ligands to
hydrogen-bond to the arginine side chain. Sterics become important because movement
of this arginine into the heme pocket to hydrogen-bond the bound-dioxygen ligand puts it
in a position to serve as a steric hindrance that stabilizes the "off" state of the enzyme.
However, it should be emphasized that this is only one possible model. Further study will
be required to elucidate the intricacies of the sensing mechanism for these FixL proteins.
Table 3.1: Statistics for met-BjFixLH data collection, and phase determination.
<table>
<thead>
<tr>
<th>Data set</th>
<th>A</th>
<th>Res. A</th>
<th>R&lt;sub&gt;sym&lt;/sub&gt; †</th>
<th>Reflections, observed/unique</th>
<th>Completeness, % ‡</th>
<th>Phasing power iso/ano</th>
<th>R&lt;sub&gt;calc&lt;/sub&gt;</th>
<th>R&lt;sub&gt;meas&lt;/sub&gt;</th>
<th>Figure of merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nat1</td>
<td>1.7396</td>
<td>2.7</td>
<td>8.1 (23.2)</td>
<td>29279/5002</td>
<td>95.3 (97.0)</td>
<td>—/2.33</td>
<td>-</td>
<td>0.014</td>
<td>0.411</td>
</tr>
<tr>
<td>Nat2</td>
<td>1.7379</td>
<td>2.9</td>
<td>6.7 (23.5)</td>
<td>18123/4098</td>
<td>95.7 (97.9)</td>
<td>—/1.08</td>
<td>-</td>
<td>0.068</td>
<td>0.254</td>
</tr>
<tr>
<td>Nat3</td>
<td>1.6984</td>
<td>2.8</td>
<td>8.0 (25.4)</td>
<td>25616/4504</td>
<td>95.2 (98.5)</td>
<td>1.02/1.64</td>
<td>-</td>
<td>0.047</td>
<td>0.351</td>
</tr>
<tr>
<td>Nat4</td>
<td>1.4938</td>
<td>2.4</td>
<td>7.6 (22.1)</td>
<td>31111/7104</td>
<td>97.0 (98.0)</td>
<td>1.38/—</td>
<td>0.664</td>
<td>0.073</td>
<td>0.265</td>
</tr>
<tr>
<td>Pt</td>
<td>1.4938</td>
<td>2.9</td>
<td>8.8 (18.4)</td>
<td>18929/3605</td>
<td>84.4 (87.4)</td>
<td>0.99/—</td>
<td>0.711</td>
<td>0.095</td>
<td>0.222</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.656</td>
</tr>
</tbody>
</table>

* $R_{sym}(I) = \Sigma_{h} \Sigma_{i} |I_{h,i}| / \Sigma_{h} I_{h}$, where $I$ is the mean intensity of the $i$ observations of reflection $h$.
† Numbers in parentheses represent the statistics for the shell comprising the outer 10% (theoretical) of the data.

Table 3.1
Table 3.2: Data processing and refinement statistics for deoxy-BjFixLH and ligand-bound complexes.
<table>
<thead>
<tr>
<th>Crystals*</th>
<th>met</th>
<th>deoxy</th>
<th>oxy</th>
<th>imidazole</th>
<th>cyanonet*</th>
<th>cyanomet*</th>
<th>MeIm</th>
<th>NO</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
<td>R32</td>
</tr>
<tr>
<td>Wavelengths (Å)</td>
<td>1.4038</td>
<td>0.9920</td>
<td>1.2830</td>
<td>1.2830</td>
<td>1.4006</td>
<td>1.0000</td>
<td>0.9184</td>
<td>1.2830</td>
<td>0.5790</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 - 2.4</td>
<td>20.0 - 2.2</td>
<td>20.0 - 2.3</td>
<td>20.0 - 2.6</td>
<td>20.0 - 2.7</td>
<td>20.0 - 2.4</td>
<td>20.0 - 2.7</td>
<td>20.0 - 2.5</td>
<td>20.0 - 2.4</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>31111</td>
<td>49059</td>
<td>57805</td>
<td>52578</td>
<td>31639</td>
<td>26220</td>
<td>25018</td>
<td>45555</td>
<td>45212</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>7104</td>
<td>9312</td>
<td>7990</td>
<td>5137</td>
<td>5150</td>
<td>7182</td>
<td>5000</td>
<td>6138</td>
<td>7132</td>
</tr>
<tr>
<td>No. of protein (ligand) atoms</td>
<td>981</td>
<td>960</td>
<td>957 (2)</td>
<td>957 (5)</td>
<td>973 (2)</td>
<td>964 (2)</td>
<td>957 (5)</td>
<td>975 (2)</td>
<td>966 (2)</td>
</tr>
<tr>
<td>B-factor of protein (Å²)</td>
<td>34.4</td>
<td>21.6</td>
<td>25.0</td>
<td>25.0</td>
<td>28.7</td>
<td>18.8</td>
<td>25.2</td>
<td>20.6</td>
<td></td>
</tr>
<tr>
<td>B-factor of ligand (Å²)</td>
<td>---</td>
<td>---</td>
<td>21.5</td>
<td>26.8</td>
<td>22.8</td>
<td>15.1</td>
<td>17.6</td>
<td>19.5</td>
<td>12.4</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>97.6 (98.0)</td>
<td>99.6 (100.0)</td>
<td>97.6 (94.9)</td>
<td>91.9 (83.3)</td>
<td>97.9 (95.3)</td>
<td>96.9 (100.0)</td>
<td>98.7 (90.7)</td>
<td>97.0 (92.3)</td>
<td>98.9 (96.6)</td>
</tr>
<tr>
<td>I / sigma</td>
<td>15.5 (7.9)</td>
<td>15.5 (7.6)</td>
<td>19.6 (10.8)</td>
<td>29.2 (16.3)</td>
<td>16.6 (7.9)</td>
<td>24.3 (7.0)</td>
<td>17.6 (6.5)</td>
<td>26.8 (12.5)</td>
<td>14.4 (6.8)</td>
</tr>
<tr>
<td>Rmerge (%)*</td>
<td>7.6 (22.1)</td>
<td>4.4 (29.1)</td>
<td>4.8 (25.6)</td>
<td>5.8 (26.0)</td>
<td>4.7 (25.8)</td>
<td>3.2 (22.1)</td>
<td>4.7 (21.7)</td>
<td>5.9 (29.3)</td>
<td>4.4 (31.5)</td>
</tr>
<tr>
<td>R factor (Rfree) (%)*</td>
<td>19.5 (24.8)</td>
<td>22.1 (26.5)</td>
<td>20.6 (25.0)</td>
<td>19.5 (26.3)</td>
<td>21.0 (24.7)</td>
<td>22.6 (26.8)</td>
<td>20.1 (22.9)</td>
<td>20.5 (25.2)</td>
<td>21.5 (25.1)</td>
</tr>
<tr>
<td>RMSD of bond distance (Å)</td>
<td>0.014</td>
<td>0.014</td>
<td>0.016</td>
<td>0.014</td>
<td>0.013</td>
<td>0.016</td>
<td>0.013</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>RMSD of bond angle (°)</td>
<td>1.68</td>
<td>1.57</td>
<td>1.80</td>
<td>1.82</td>
<td>1.62</td>
<td>1.73</td>
<td>1.55</td>
<td>1.78</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Table 3.2: Data processing and refinement statistics for deoxy-BjFixLH and ligand-bound complexes. (continued)
The CO-BjFixLH data were collected on beamline 9-2 at SSRL. The met-, oxy-, imidazole-, NO- and cyanomet-BjFixLH data were collected on X4A at the National Synchrotron Light Source in Brookhaven National Laboratory. The Melm-BjFixLH data were collected on BIOCARS 14D at the Advance Photon Source at Argonne National Laboratory. The deoxy-BjFixLH data were collected on beamline 9-1 at SSRL.

The crystal was prepared by soaking the met-FixLH crystal with 100mM potassium cyanide.

The crystal was grown in synthetic mother liquor containing 10mM potassium cyanide.

The numbers in parentheses are for the highest resolution shell.

\[ R_{\text{merge}}(I) = \frac{\sum_h \sum_i |I_i - I|}{\sum_h \sum_i I}, \text{ where } I \text{ is the mean intensity of the } i \text{ observations of reflection } h. \]

\[ R \text{ factor} = 100 \times \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}, \text{ where } F_{\text{obs}} \text{ and } F_{\text{calc}} \text{ are the observed and calculated structure factors, respectively. } R_{\text{free}} \text{ was calculated using } 8\% \text{ of reflections.} \]

Table 3.2 Continued.
Figure 3.1: Crystals of met-BjFixLH.
Figure 3.2: Harker sections of the Patterson maps of met-BjFixLH: (A) anomalous Patterson map using data from Nat3; (B) isomorphic Patterson map using data from Nat1 and Nat3.
Figure 3.3: Ribbon diagrams of deoxy-BjFixLH. The secondary structure elements are color-coded with α-helices as red, β-sheets as cyan, and random coils as green. The atoms of the heme cofactor and the proximal histidine are shown as ball-and-stick models and are colored by their elements, with carbon as gray, nitrogen as blue, oxygen as red, and iron as green.
Figure 3.4: Structure-based sequence alignment of BjFixL, RmFixL, PYP, HERG and LOV2 PAS domains. Helices are represented by rectangular boxes, strands by arrows and loops by lines. The striped region at the beginning of F-helix varies depending on the length of the helix. The amino acids labeled in red are active site residues that interact with cofactors. Conserved hydrophobic residues, Gly/Pro residues, and conserved residues whose side chains form hydrogen bonds to main chain atoms are highlighted in green, yellow and blue, respectively.
Figure 3.5: The conserved PAS fold. (A) Overlap of PAS domain regions of PYP (yellow), HERG (blue), BjFixLH (red), and LOV2 (green). (B) *E. halophila* PYP colored according to the original PYP nomenclature (110): N-terminal cap, tan; PAS core, violet, helical connector, green; and β-scaffold, red.
Figure 3.6: (A) The structure of the heme pocket in deoxy-BjFixLH. The hydrogen bonds are shown as dashed lines; (B) Corresponding 2F_o-F_C electron density map (1σ).
Figure 3.7: Ribbon and ball-and-stick model diagrams of the BjFixLH heme-binding pocket for (A) met-BjFixLH, (B) deoxy-BjFixLH, (C) CO-BjFixLH, (D) NO-BjFixLH, (E) oxy-BjFixLH, (F) the co-crystallized cyanomet-BjFixLH, (G) imidazole-BjFixLH, and (H) MeIm-BjFixLH. W1 represents water. The dashed lines show the hydrogen bonding or the salt-bridge interactions.
Figure 3.8: Ball-and-stick diagrams of three heme-binding pockets. Structures are shown for *Glycera dibranchiata* hemoglobin (PDB ID: 2HBG) (left), BjFixLH (center), and the NO transporter protein (PDB ID: 1NP1) (right) (116, 117). The rightmost and leftmost side chains correspond to the E7 and E11 residues of hemoglobins, respectively. The additional side chain in BjFixLH (colored in red) corresponds to Ile 215. The structures were aligned based on the orientation of the proximal histidine and the porphyrin ring. The atoms are colored as in Figure 3.1.
Figure 3.9: Overlap of the structures of deoxy-BjFixLH (blue) and met-BjFixLH (tan).
Figure 3.10: Plots of RMSD following least square fitting of deoxy- and each of the six
ligand-bound structures (CO-, NO-, cyanomet-, imidazole-, MeIm-, and oxy-BjFixLH) to
met-BjFixLH: (A) The RMSD for each residue based on its main-chain atoms and (B) the
overall RMSD for each structure based on all the atoms, the main-chain atoms, and the
heme.
A. Main-chain RMSD for each residue with respect to met-BjFixLH

B. Overall RMSD with respect to met-BjFixLH

Figure 3.10
Figure 3.11: (A) Ribbon diagrams of oxy-BjFixLH colored by secondary structure (α-helix, red; β-strand, blue; random coil, yellow; the FG loop, green). The atoms of the heme, dioxygen ligand and proximal histidine are depicted as ball-and-stick models and are colored by element (carbon, gray; oxygen, red; nitrogen, blue; and iron, green). (B) $F_O-F_C$ difference map (4.0σ) of the oxy-BjFixLH heme pocket with the dioxygen ligand omitted.
Figure 3.12: Conformation differences in the FG loop (A) and in the heme pocket (B) between deoxy-BjFixLH (cyan and blue) and oxy-BjFixLH (red and tan). The structure of the heme pocket in oxy-BjFixLH (C) and the corresponding 2F_{o}-F_{c} electron density map (1σ). The hydrogen bonds are shown as dashed lines.
Figure 3.13. Structural comparison of the heme environment for (A), sperm whale oxy-myoglobin; and (B), oxy-BjFixLH.
Figure 3.14: (A) Ribbon diagram of CO-BjFixLH colored by secondary structure (α-helix, red; β-strand, blue; random coil, tan; the FG loop, green). The atoms of the heme, dioxygen ligand and proximal histidine are depicted as ball-and-stick models and are colored by element (carbon, gray; oxygen, red; nitrogen, blue; and iron, green). Ball-and-stick models with electron density, for the BjFixLH heme binding pocket of (B) CO-BjFixLH, and (C) the co-crystallized cyanomet-BjFixLH (2F_O-F_C density (blue, 1.0 σ) and F_O-F_C omit density (red, 3.0 σ)).
Figure 3.15: Proposed mechanism for the heme-driven conformational change of BjFixLH. The heme-propionate 6 and 7 are labeled HP6 and HP7, respectively. Arg 220 and Arg 206 are colored in blue, and dioxygen in red.
CHAPTER 4

Crystal structure of a methanogen methyltransferase from *Methanosarcina barkeri*: identification of a new UAG-encoded amino acid

4.1 Introduction

Methanogens are an unusual class of archaebacteria that produce methane as part of their anaerobic growth (132). They are typically found in oxygen-deficient environments, such as marshes, swamps, sludge (formed during sewage treatment), and the digestive systems of ruminants. These methanogens are the most evolutionally primitive living organisms known to date and are implicated in the origin of life (133, 134). They are also extremely important for the digestion of nonabsorbable fermentable materials in higher organisms (including humans), and the degradation of pollutants in current environments (135, 136).

All methanogens can obtain their energy by reducing carbon dioxide and oxidizing hydrogen to methane and water: \( \text{CO}_2 + 4\text{H}_2 \rightarrow \text{CH}_4 + 2\text{H}_2\text{O} \). Some methanogens can grow on other substrates as well (132). For example, the methanogenic organism, *Methanosarcina barkeri*, can utilize almost all of the known methylotrophic methanogenic substrates, including methanol, acetate, methylated sulfur compounds, and methylamines (137-140). One of the pathways involves the activation and transfer of the methyl group from methylamines. The catabolism of methylamines by methanogens...
involves a conserved arrangement of proteins (141) (Figure 4.1). A specific monomethylamine (MMA), dimethylamine (DMA), or trimethylamine (TMA) methyltransferase activates the substrate for methyl transfer to a cognate corrinoid protein (138, 139, 142). A second methyltransferase catalyzes the transfer of the methyl group from the methylated corrinoid cofactor to the thiol of 2-mercaptoethanesulfonic acid (coenzyme M, or CoM) forming methyl-CoM (140, 143). Methyl-CoM is subsequently used to generate methane by methyl-CoM reductase (141).

An unusual feature conserved in all known methylamine (MMA, DMA, or TMA) methyltransferase genes is the presence of a single in-frame amber (UAG) codon that does not stop translation during protein synthesis (137, 144) (Figure 4.1). Such conservation is intriguing in light of the low sequence similarity among different methylamine methyltransferase proteins. Originally this internal amber codon was found in the different *Methanosarcina* species, including *M. barkeri* strains MS, NIH and Fusaro, *M. thermophila*, and *M. mazei* (137, 144, 145). Recently, however, a homolog has been found in the genome of the Gram-positive bacterium *Desulfotobacterium hafniense* that also contains a single in-frame amber codon (146).

Several mechanisms have been identified and developed to suppress the stop codon in both prokaryotes and eukaryotes (147, 148). In some cases, for example, the stop codon is bypassed by frameshifts, or by the presences of multiple copies of the gene in the genome ((149, 150). The sequence analyses in methylamine methyltransferases, however, have eliminated the possibility of circumventing the amber codon by one of these means (137, 144). For example, the other reading frames of MMA, DMA and TMA methyltransferase genes produce numerous UAA and UGA stops, which cannot be used
to synthesis the proteins with the right size. Moreover, three copies of the DMA methyltransferase genes have been found in the genome of *M. barkeri*, but all of them contain a single in-frame UAG codon in the same position. In addition, only a single TMA methyltransferase gene that contains the UAG codon was found.

Previously, analysis of tryptic fragments of MMA methyltransferase (MtmB) by mass spectrometry and Edman degradation suggested that the amber codon serves as a sense codon that corresponds to lysine (145). The harsh conditions of peptide isolation, however, left open the possibility of a modification on the lysine residue signaled by the amber codon. Such rare use of what is normally a stop codon to signal incorporation of an unusual amino acid has precedent in the use of the opal codon, UGA, to encode selenocysteine, the twenty-first amino acid to be found in Bacterial, Eucaryal, and Archaeal proteins (151, 152).

Another important issue related to this system is how these methylamine methyltransferases transfer the methyl group to their cognate corrinoid proteins. This reaction is the starting point of the methylamine-dependent CoM methylation. Monomethylamine catabolism in *M. barkeri*, for example, is mediated by a monomethylamine methyltransferase, MtmB, that assists in activating the methyl group of the monomethylamine substrate so that it can be transferred to an associated corrinoid protein, MtmC. Both of the proteins have been purified to homogeneity by Dr. Krzycki’s group (139). During their studies of these proteins, they found that MtmB and MtmC form a stable complex (MtmBC) when isolated aerobically. Since MtmB and MtmC must interact with each other in order to transfer the methyl group, the MtmBC is an excellent
system for studying the interactions and potential conformational changes that take place during the methyl transfer reaction.

In an effort to clarify the identity of the UAG-encoded residue and to understand the mechanism of methyl transfer reactions in methanogens, the structure of the *M. barkeri* MS monomethylamine methyltransferase (MtmB) was elucidated. Two forms of the enzyme were determined to 1.55 Å and 1.7 Å resolution, respectively. Both structures suggest that the amber-encoded residue is distinct from the other 21 natural amino acids. Moreover, crystals of *M. barkeri* MtmBC complex have also been obtained and a 6.6 Å dataset has been collected. Many efforts have been made to improve the diffraction quality of these complex crystals.

4.2 Materials and methods

4.2.1 Crystallization and data collection

Purified MtmBC complex was provided by the Krzycki’s laboratory. The protein was isolated according the procedure previously described for MtmB with the exception that the procedure was performed aerobically (139). Under these conditions, MtmB forms a relatively stable complex with MtmC. Dynamic light scattering experiments of the protein sample by our lab indicated that the protein was monodispersive with an average molecular mass around 257 kDa (Figure 4.2). This would correspond to a trimer of MtmBC complex (the MWs of the MtmB and MtmC monomer are about 50 and 23 kDa, respectively). Crystallization attempts on the MtmBC complex were performed at 4 °C using the hanging-drop vapor-diffusion method (65). Drops contained equal volumes of
reservoir and protein solution (30 mg/mL protein, 200 mM NaCl, and 50 mM MOPS, pH 6.0).

**Crystallization and data collection of MtmB** Clear crystals of different shapes could be obtained with several different crystallization conditions from crystal screen kits I and II (Hampton Research). Two of these were used in the structural analysis {form 1: 4.3 M NaCl, and 0.1 M HEPES (pH 7.5); and form 2: 1.6 M (NH₄)₂SO₄, 0.1 M NaCl and 0.1 M HEPES (pH 7.5)} (Figure 4.3A). SDS-PAGE gels of washed and redissolved crystals revealed that they contained only MtmB (Figure 4.3B). Prior to data collection, the MtmB crystals were transferred gradually into the reservoir solution containing increasing amounts of glycerol up to 30% (v/v). The two high-resolution native data sets were collected at the BIOCARS sector 14C of the Advance Photon Source (APS) at Argonne National Laboratory. The data for the heavy-atom derivatives were collected at beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL), BIOCARS sector 14D at APS, and beamline X4A at the National Synchrotron Light Source in Brookhaven National Lab. For each heavy-atom derivative, anomalous diffraction data were collected at the wavelength corresponding to the peak anomalous scattering, except for iodine-and cesium-containing derivatives, which were collected at low energy to optimize the anomalous signal. Data processing and reduction were performed using DENZO and SCALEPACK (66). The MtmB crystals contain one MtmB subunit per asymmetric unit and belong to space group P6₃22 with unit cell dimensions a = b = 158.8 Å, c = 136.5 Å. The estimated solvent content is about 75%.

**Crystallization and data collection of the MtmBC complex** Crystals of the MtmBC complex were obtained using the same sample as for the MtmB crystals. The only
difference was the crystallization condition used. The well solution contained 2 M (NH₄)₂SO₄ and 5% isopropanol. The crystals appeared pink consistent with the existence of the corrinoid cofactor (Figure 4.3C). SDS-PAGE gels of washed and redissolved crystals revealed that they contained both MtmB and MtmC proteins (Figure 4.3D). These crystals are very difficult to obtain diffraction from. The current procedure to improve the resolution involves crosslinking the crystals with 25% glutaraldehyde for three hours by vapor diffusion in original crystallization drop, and then transferring the crystals stepwise to the well solution containing decreasing concentration of isopropanol over the course of a week. The crystals were then washed briefly with saturated Li₂SO₄ solution for a couple of minutes before flash freezing in liquid N₂. So far, the best crystals have diffracted to about 6.6 Å at the BIOCARS sector 14B of the APS at Argonne National Laboratory. These diffraction data were processed, integrated, and scaled with DENZO and SCALEPACK (66). The complex crystals have space group R32 with the cell parameters a = b = 297.8 Å, c = 186.1 Å. The calculated Matthews coefficient V_M of 2.7 Å³Da⁻¹ suggests four complex molecules (MW = 73,232 Da per MtmB and MtmC) per asymmetric unit and a solvent content of 53% (153). Due to the ambiguity caused by the low resolution, the diffraction data from the complex crystals could also be processed with the same space group R32, but with a smaller cell, a = b = 151.0 Å, c = 179.3 Å. In this cell, there is one complex molecule per asymmetric unit corresponding to the Matthews coefficient V_M of 2.7 Å³Da⁻¹.
4.2.2 Structure determination and refinement

Structure determination and refinement for MtmB The structure of MtmB was determined by the multiple isomorphous replacement and anomalous scattering (MIRAS) method. The derivatives were prepared by soaking the native crystals with heavy-atom compounds, including NaI, I₂, EMTS (ethylmercurithiosalicylic acid), CH₃PbOAc, K₂OsO₄, and CsCl (Table 4.1). Six data sets were used in final phase calculation.

The Hg and Pb sites were independently located by isomorphous and anomalous difference Patterson maps (Figure 4.4). All the other derivatives were solved by difference Fourier using the Hg/Pb phases. The best derivative was obtained with sodium iodide (Table 4.1). Initially heavy atom parameters were refined and MIRAS phases were calculated and compared using PHASES (103) and MLPHARE (154) from CCP4 program suite (155). These phases were further improved by solvent flattering and histogram matching with DM (156) and maximum likelihood density modification with RESOLVE (157, 158). Model building was carried out with the program O (68, 104). Despite the weak anomalous signals and common sites, at this stage, over 300 residues of polyAla model could be built into this map. Phase calculated from this preliminary model was recombined with the experimental phases using PHASES (103), CNS (69) and SIGMAA (159), respectively. The density map calculated from the recombined phase was used to adjust and complete the backbone model. The secondary structure prediction from the MtmB amino acid sequence by GOR IV (160) was used as an aid to confirm the correctness of the trace. Multiple cycles of such model building and phase improvement were carried out until over 97% of visible side chains could be built. Structure refinement was performed using CNS (69) with 10% of the data omitted for the Free-R factor
calculation. No sigma cutoff was used. The starting R and Free-R factors after the first rigid body refinement were 0.45 and 0.44. The model was then refined by simulated annealing, followed by several cycles of minimization using a maximum likelihood target based on the amplitudes and incorporation of a flat bulk solvent correction and an overall anisotropic B-factor. This refinement dramatically decreased the R/Free-R factors to lower than 20% (Table 4.2). The quality of the final model of the MtmB was assessed using the program PROCHECK and was found to be acceptable (71). The statistics for each of the stereochemical parameters was inside or better than the expected values, and 90% of the residues being in the most favored regions of the Ramachandran plot. The refined model consists of all the residues from 2 to 458, except the N-terminal methionine. This is consistent with its posttranslational cleavage as determined by N-terminal sequencing of the isolated protein (137). There are a total three residues having the cis conformation, Pro 204, Glu 235 and Val 367.

The figures were prepared with the programs XtalView (76), MolScript (74), Raster3D (75), BobScript (161), and Grasp (162).

**Structure determination of the MtmBC complex** Molecular replacement solution for MtmBC complex structure was searched using the programs CNS (69), Amore (163), Molrep (164), GLRF (165) and EPMR (166). Both MtmB and the cobalamin-binding domain of methionine synthase (MetH) (PDB ID 1BMT) structures were used as the search model.
4.3 Results and discussions

4.3.1 The protein fold of MtmB

The structures of MtmB were solved using two forms that differed only in the precipitating salt used (NaCl for form 1, and (NH₄)₂SO₄ for form 2) for crystallization. The overall structures of these two forms are virtually identical and consist of a homohexamer arranged into a dimer of trimers with overall D3 symmetry (Figure 4.5). Each subunit adopts an α/β TIM barrel fold (167, 168) (Figure 4.6A) that is reminiscent of other recently determined corrinoid-cofactor containing proteins: tetrahydrofolate:corrinoid/ironsulfur protein methyltransferase (169), diol dehydratase (170), and racemases/mutases (171, 172). No significant sequence similarity, however, is found between MtmB and these proteins. The secondary structure assignment using PROCHECK (71) are: α helixes: α1, A9-A18; α2, A24-A42; α3, A56-A73; α4, A88-A95; α5, A137-A148; α6, A174-A193; α7, A210-A214; α8, A240-A251; α9, A271-A288; α10, A309-A325; α11, A343-A359; α12, A380-A393; α13, A398-A412; α14, A438-A453; β strands: β1, A19-A21; β2, A75-A77; β3, A82-A87; β4, A101-A104; β5, A111-A114; β6, A125-A135; β7, A155--A161; β8, A201-A203; β9, A227-A231; β10, A255-A264; β11, A293-A300; β12, A330-A336; β13, A363-A368; β14, A428-A430; β15, A434-A436. The eight-stranded beta barrel is formed from strands β6 to β13 and is surrounded by eight major α helices α5-α12. Consistent with other methyltransferases, this barrel forms a deep cavity that for the MtmB is negatively charged (Figure 4.6B). This may facilitate binding of the methylammonium cation. The amber-encoded residue lies precisely at the bottom of this cavity suggestive of a potential role in catalysis.
4.3.2 The UAG-encoded residue of MtmB

Previous biochemical analyses of the trypsin proteolysed fragments of MtmB indicated that the residue encoded by the amber codon of the *mtmB* gene is a lysine or a lysine-derivative (145). The structures from both NaCl and (NH₄)₂SO₄ crystal forms support the latter assignment with clear 3σ 2Fo-Fc density for a lysine modified at its epsilon nitrogen by a substituted pyrroline ring (Figure 4.7), which is distinct from any of the 21 known natural amino acids. Based on our analysis, the identity of the modifying group appears to be (4R, 5R)-4-substituted-pyrroline-5-carboxylate, with the carboxylate of the modifying group in amide linkage to the lysine epsilon nitrogen (Figure 4.8 and 4.9B). We propose that this UAG encoded amino acid be named L-pyrrolysine.

4.3.2.1 Structure of the UAG-encoded residue in the absence of ammonium ion

The initial model of the UAG-encoded residue was derived from the 1.55 Å resolution structure of the NaCl crystal form (Figure 4.7A). In this crystal form, the initial 3σ 2Fo-Fc density for UAG-encoded residue could be fit to β-methyl-D-proline in amide linkage with the epsilon nitrogen of lysine (orientation 1) (Figure 4.8). Position 1 in the five-membered ring was modeled as a nitrogen atom based on the presence of hydrogen bonding interactions from Glu 259 and Glu 229 (Figure 4.11A). The substituent at position 4 currently remains unidentified. Based on the current fit to the electron density, it could be a methyl, ammonium, or hydroxyl group. It is within hydrogen-bonding distance of Tyr 335, but this distance is long, 3.16 Å, and two waters hydrogen bond to Tyr 335 with a better geometry.
After fitting and refinement of the initial model, weak but broad $3\sigma$ difference density appeared. This density was best accounted for by the presence of a second orientation of the UAG-encoded residue (orientation 2), though at much lower occupancy (Figure 4.9). In this second orientation, the ring is rotated approximately $90^\circ$ relative to its position in the orientation 1. Refinements of the occupancies and thermal parameters of both orientations suggest an occupancy for orientation 2 of only 15% in the NaCl crystal form, while orientation 1 has an occupancy of 85%. Orientation 2 is of significance, however, since this orientation is the major orientation observed in the (NH$_4$)$_2$SO$_4$ crystal form.

### 4.3.2.2 Structure of the UAG-encoded residue in the presence of ammonium ion

While the overall structure of MtmB for the (NH$_4$)$_2$SO$_4$ crystal form at 1.7 Å resolution is virtually identical to that for the NaCl crystal form, the electron density around the UAG-encoded residue is more complex. The 2F$_{\text{o}}$-F$_{\text{c}}$ density of the pyrroline ring of the UAG encoded residue in the (NH$_4$)$_2$SO$_4$ crystal form differs significantly from that observed in the NaCl crystal form (Figure 4.7B). This density could eventually be fit, however, to a model comprised of the same two orientations of the pyrroline ring as found in the NaCl crystal form, but at different occupancies (Figure 4.10). While the orientation 1 was the dominant conformation in the NaCl crystal form, in the (NH$_4$)$_2$SO$_4$ crystal form, the relative occupancies of orientations 1 and 2 are 40% and 60%, respectively. Thus, orientation 2 now becomes the dominant conformation.

In addition to the difference in the relative occupancies, electron density maps of the (NH$_4$)$_2$SO$_4$ crystal form also reveal the presence of an additional atom in both
orientations (Figure 4.10). In orientation 2, the additional density is attached at the C-2 ring carbon and is assigned to a nitrogen atom since (NH₄)₂SO₄ is the only species not present in the solution used to grow the NaCl crystals. Consistent with this assignment, the side chains of Glu 259 and Gln 333 are at hydrogen bonding distances to the proposed amine substituent (Figure 4.11C). These interactions likely stabilize the L-pyrolysine in orientation 2, thereby accounting for its higher occupancy in the (NH₄)₂SO₄ crystal form. In order to remove the remaining difference density, a free ammonium ion was modeled in the orientation 1 model at 40% occupancy (Figure 4.10 and 4.11B). This ion does not bind or interact with the UAG-encoded residue, but instead hydrogen bonds to Met 261 and Tyr 335. The ammonium ion occupies a position near the amine group bound to the 5-membered ring in orientation 2 of the UAG-encoded residue in the (NH₄)₂SO₄ crystal form, and may reflect an intermediate state prior to amine addition.

Chemically, the addition of the presumed amine has implications regarding the features of the ring. In order for an amine to add to the ring, the C-2 carbon must be $sp^2$ hybridized. This most likely occurs because of the presence of a double bond between the N-1 and C-2 atoms of the ring. This implies that the identity of the lysine-modifying group before amine addition is (4R, 5R)-4-substituted-pyrroline-5-carboxylate in amide linkage to the epsilon nitrogen of lysine (Figure 4.9B). After amine addition, the identity of the modifying group is (2R, 3R, 5R)-5-amino-3-substituted-pyrroline-2-carboxylate in amide linkage to lysine (Figure 4.10B).
4.3.2.3 Mass spectroscopic support for interpretation of the UAG-encoded residue

(Done by Carey M. James from Dr. Krzycki’s group)

As the presence of pyrrolysine should increase the molecular weight of the MtmB in a predictable way, the MtmBC complex was subjected to electrospray mass spectrometry to confirm its assignment. The measured masses of MtmB and MtmC were 50,105 ± 2 Da and 23,066 ± 1 Daltons, respectively. These values can be compared to the deduced mass from the predicted protein sequences from the encoding genes (137) using the program SHERPA (173). Without the corrinoid prosthetic group, the theoretical average mass of MtmC is calculated as 23,067 Da, which is consistent with the experimentally derived molecular mass. For MtmB, the theoretical average mass is calculated as 49,998 Da, assuming a lysine residue at the UAG-encoded position. This calculated value is 107 atomic mass unit (amu) less than the experimental mass of MtmB (50,105 Da). With the incorporation of the 4-pyrroline-5-carboxylate group as found in the crystal structure, however, the theoretical molecular weight for MtmB increases by 109 amu to 50,107 Daltons assuming that a methyl group is attached to the C-4 ring carbon. If the substituent attached to the ring is either an amine or hydroxyl group, then the theoretical mass would be higher by 110 amu (50,108 Da) and 111 amu (50,109 Da), respectively. These mass spectra data support our current assignment of the UAG-encoded amino acid.

4.3.3 A proposed mechanism for monomethylamine activation by pyrrolysine

As mentioned previously, the location of the amber-encoded amino acid, pyrrolysine, suggested a potential role in catalysis. A fundamental question, therefore, is
how pyrrolysine activates methylamine substrate and facilitates the transfer of the methyl group to the corrinoid cofactor of MtmC. One possibility for how this could be achieved is based on our two different forms of the enzyme (Figure 4.12). In this model, the role of pyrrolysine is to position and activate the methyl group of monomethylamine for transfer to the corrinoid cofactor.

The three major states of pyrrolysine in the two crystal forms exhibit different interactions with the surrounding side chains (Figure 4.11). By aligning these structures in sequence, one can obtain a picture of the events taking place in MtmB. The first structure (NaCl form, orientation 1) provides a snapshot of the protein with no substrate added (Figure 4.11A). The second structure ((NH₄)₂SO₄ form, orientation 1) shows the initial binding of amine substrates to the protein (Figure 4.11B). Here Gln 333 and Tyr 335 serve as hydrogen-bonding acceptors stabilizing amine binding. A more important residue is Glu 259. Prior to methylamine addition, the carboxylate of Glu 259 is probably protonated and positioned so that this proton hydrogen-bonds the pyrroline ring nitrogen of pyrrolysine. This feature is important since protonation of the Schiff base nitrogen makes it more electron withdrawing, and this in turn helps to activate the C-2 carbon for nucleophilic addition (Figure 4.14). Following methylamine addition to the C-2 carbon ((NH₄)₂SO₄ form, orientation 2), the new deprotonated carboxylate of Glu 259 shifts to be a hydrogen-bond acceptor with the bound methyl ammonium group (Figure 4.11C). The amide side chain of Gln 333 provides a second hydrogen-bond acceptor to this bound methyl ammonium group. Importantly, these two hydrogen bonds and the covalent bond to the pyrroline ring not only help to stabilize the position change of methyl ammonium substrate making methyl transfer more favorable, but also help to position the methyl
ammonium group so that its methyl group is directed towards the surface of the binding cleft. Here, it presumably is positioned to interact with the corrinoid-cofactor of MtmC upon formation of the MtmBC complex.

4.3.4 Docking model of MtmB/MtmC

In order for methyl group to be transferred from MtmB to MtmC, these proteins must come together in a fashion this is favorable for the methyl transfer reaction. Recently, structure of a C-terminal fragment of B12-dependent methionine synthase (MetH) has been determined (174). This fragment contains both cobalamin- and substrate-binding domains. Comparison with the structures of the individual domains, a major domain rearrangement occurred in the “cap” region of the cobalamin-binding domain that presumably facilitates the formation of a new interface between cobalamin and the substrate-binding domain. However, there is no conformational change observed upon binding of the substrate (methyl donor) in the relatively low-resolution structure (3.8 Å), and the substrate is about 6 Å away from the corrinoid ring. It remains unclear how the methyl group is transferred.

Since the cobalamin-binding domain of MetH has 32% sequence identity and 53% similarity with MtmC, these proteins presumably have a similar fold and structure (137). A preliminary docking model was prepared between MtmB and the cobalamin-binding domain of MetH (PDB ID: 1K7Y). First, the MtmB was orientated to the similar position as the substrate-binding domain of MetH. Then, the cobalamin-binding region of MetH was docked into the MtmB based both on its shape and the position of the bound cobalamin-cofactor in the superimposed diol dehydratase structure (PDB ID: 1EGM)
(Figure 4.13). One important result from the docking model was that the position of the cobalamin-cofactor is optimally positioned to abstract the methyl group of the activated methylamine, supporting the role of the UAG-encoded residue in methylamine activation.

4.3.5 Preliminary results for the crystallization and structural studies of the MtmBC complex

4.3.5.1 Improvement of the quality of the crystals

While the docking model of MtmB/MtmC provides one possible mechanism that how the methyl group could be transferred from the methylamine substrate to the corrinoid cofactor through the interaction with pyrrolysine, direct insights could be obtained from structure of the MtmBC complex. From the initial crystallization screens, large crystals of the MtmBC complex were obtained from the condition containing 2.0 M (NH₄)₂SO₄ and 5% (V/V) isopropanol. X-ray diffraction of these crystals, however, was initially undetectable. Hence, much effort has been devoted to optimize the crystallization conditions and to stabilize the crystals. Several ways have been tried to improve the quality of the crystals:

Organic solvent screen: It has been found that the presence of 2-5% isopropanol is essential to obtain the MtmBC crystals. Due to the high volatility of isopropanol, however, these crystals are extremely sensitive and difficult to handle. Thus, a series of organic solvents, including glycerol, butanol, ethylene glycol, acetonitrile, ethanol, dioxane, MPD, and PEGs, have been used to substitute isopropanol in the crystallization condition. Also, the amount of isopropanol in the crystal has been gradually changed to make it possible to transfer and mount crystal.
Additive screen: Various additives have been either added to the crystallization condition or treated with the crystals to improve the stability of the crystals. These additives include heavy-atom agents, precipitating salts, detergents, enzyme substrates, and inhibitors.

Cross-linking agent screen: Different types of cross-linking reagents have been used to stabilize the crystals before handling, such as glutaraldehyde (amine linkage), DMP (amine linkage), EDC (carboxyl-amine linkage), BM(PEO)$_3$ (sulfhydryl linkage), and SIA (sulfhydryl-amine linkage).

Cryoprotectant screen: Among a variety of conventional cryoprotectants, including glycerol, paratone-N, PEG400, MPD and cryo-salts, so far, the best one for MtmBC complex crystals is saturated lithium sulfate.

Taken together, removal of isopropanol, followed by cross-linking with glutaraldehyde, and suitable cryoprotectant lithium sulfate, enabled the resolution of the MtmBC complex data up to 6.5 Å.

4.3.5.2 Structural analysis by molecular replacement (MR) method

Since there are no MAD or heavy-atom derivative data available to this point, we have attempted to solve the structure of MtmBC complex by molecular replacement method. Theoretically, two structures could serve as the potential search models for MR. One is MtmB that occupies about 68% population of the complex. Another one is the cobalamin-binding domain of MetH that shares about 32% sequence identity with MtmC. Several MR programs, including CNS, Amore, Molrep, GLRF and EPMR, have been tried preliminarily. So far, however, all of them failed. This is possibly due to the
combination of two factors. First, as mentioned above, the crystals often cracked during
the mounting procedures and the resolution of our best data is only around 7 Å. Hence,
interpretation of the space group of the crystals would have been difficult, and the
estimating errors (R-factor) of the data sets are fairly high comparing to a normal
standard. The lack of an accurate data and correct assignment for the space group
probably account for the major reasons for the MR searching failure; Second, both of the
homologous models represent only a limited portion of the complex structure, which in
practice would not be sufficient to find a right MR solution.

4.4 Conclusion

In conclusion, our structural data suggests that the in-frame amber codon in mtmB
genes encodes a novel amino acid that may play a role in catalysis of methyltransferase
reactions involving amines. Furthermore, an unusual aminoacyl-tRNA synthetase and a
suppressor tRNA_{CUA} are encoded near the mtmB genes in M. barkeri (146). While it
formally remains possible that modification of lysine occurs following UAG directed
insertion of lysine into the protein, the use of a canonical stop codon and dedicated tRNA
is most consistent with the direct translational encoding of L-pyrrolysine. Taken together,
current evidence supports the idea that L-pyrrolysine represents the 22\textsuperscript{nd} naturally
occurring amino acid to be identified in nature.
Figure 4.1: Conserved pathways for methylotrophic growth. The names of the different methyltransferases and corrinoid proteins are indicated in the boxes. Each methylamine methyltransferase is labeled with a UAG indicating that it contains an in-frame amber codon. The shaded proteins are predicted based on sequence similarity, but have not been demonstrated biochemically.
Figure 4.2: Dynamic light scattering measurement of the MtmBC complex (baseline-1.002). The narrow distribution of hydrodynamic radius suggested that MtmBC complex is monodispersive in the solution.
Figure 4.3: Crystals of (A) the *M. barkeri* monomethylamine methyltransferase (MtmB), and (C) the complex of MtmB and its conjugated corrinoid protein MtmC (MtmBC). The SDS-polyacrylamide gel of the dissolved crystals of (B) MtmB and (D) MtmBC complex. Lane 1, MW marker; Lane 2, washed and dissolved crystals.
Figure 4.4: Harker sections of the isomorphous (A) and anomalous (B) difference Patterson maps of the Hg derivative (Table 4.1). A single peak at the same location in both maps was observed. Heights of the Harker peaks in both maps were 5\( \sigma \).
Figure 4.5: Ribbon diagram of the MtmB hexamer. The subunits forming the two trimers are shaded in red and blue hues, respectively.
Figure 4.6: Structure and folding of an MtmB subunit. (A) Ribbon diagram of one MtmB subunit. The secondary structure elements are color-coded with α-helices as red, β-sheets as blue, and random coils as green. The atoms of the UAG encoded residue are shown as ball-and-stick models and are colored by their elements, with carbon as gray, nitrogen as blue, and oxygen as red. (B) Electrostatic potential energy surface diagram of MtmB in a similar orientation.
Figure 4.7: 2F₀-Fᶜ electron density omit map of the UAG-encoded residue for crystals grown in (A) NaCl solution and (B) (NH₄)₂SO₄ solution.
Figure 4.8: Current fit of (4R, 5R)-4-substituted-pyrroline 5-carboxylate to the 2F$_{O}$-F$_{C}$ electron density of NaCl crystal form.
Figure 4.9: Electron density maps of the NaCl crystal form (A) 2F_0-F_C omit map at 4.5\(\sigma\) for the 4-methyl-pyrroline carboxylate group of the UAG-encoded residue for crystals grown in NaCl solution with the two ring orientations used to fit it. The carbons of orientation 1 are colored in pink. In orientation 2 they are colored in tan. (B) Stick-diagram of proposed L-pyrolysine amino acid (orientation 1 in NaCl crystal form). (C) F_0-F_C difference map at 4.5\(\sigma\) for NaCl crystals following incorporation of model for orientation 2 at 15% occupancy. (D) Fit of orientation 1 model to this density. (E) F_0-F_C difference map at 4.5\(\sigma\) for NaCl crystals following incorporation of model for orientation 1 at 85% occupancy. (F) Fit of orientation 2 model to this density.
Figure 4.9
Figure 4.10: Electron density maps of the (NH₄)₂SO₄ crystal form (A) 2F₀-Fₐ omission map at 4.5σ for the 4-methyl-pyrroline carboxylate group of the UAG-encoded residue for crystals grown in high (NH₄)₂SO₄ solution with the two ring orientations used to fit it. The carbons of orientation 1 are colored in pink. In orientation 2 they are colored in tan. (B) Stick-diagram of proposed L-pyrolysine amino acid after amine addition (orientation 2 in (NH₄)₂SO₄ crystal form). (C) F₀-Fₐ difference map at 4.5σ for (NH₄)₂SO₄ crystals following incorporation of model for orientation 2 at 60% occupancy. (D) Fit of orientation 1 model to this density with additional ammonium ion. (E) F₀-Fₐ difference map at 4.5σ for (NH₄)₂SO₄ crystals following incorporation of model for orientation 1 at 60% occupancy. (F) Fit of orientation 2 model to this density with additional amine group added to the pyrroline ring.
Figure 4.10
Figure 4.11: Stereoview of primary forms of the active site around the amber-encoded amino acid: (A) NaCl crystals; (B) (NH$_4$)$_2$SO$_4$ crystals, 40% occupancy orientation that is similar to NaCl crystals; (C) (NH$_4$)$_2$SO$_4$ crystals, 60% occupancy orientation with amine added to ring.
Figure 4.12: Hypothetical model for the role of the amber-encoded residue in catalysis. The proposed intermediates for (a), (b), (c), and (f) are based on the structures of L-pyrolysine (X = Me, NH₂, or OH) in its 85% occupancy orientation in NaCl crystal form and on both orientations of L-pyrolysine in the (NH₄)₂SO₄ crystal form. Intermediates (d) and (e) are based on a preliminary docking model of MtmB with its cognate corrinoid protein, MtmC.
Figure 4.13: Overview of the docking model between MtmB and the cobalamin-binding domain of MetH (PDB ID: 1K7Y). (A) Molecular surface of MtmB and the stick diagram of the cobalamin-binding domain of MetH. The cobalamin cofactor is colored red. (B) Putative active site of the docking model. The main-chain atoms of the cobalamin-binding domain and MtmB are colored as green and grey, respectively. The cobalamin cofactor and the UAG-encoded amino acid (orientation 2 in (NH₄)₂SO₄ crystal form) are colored by element, with nitrogen as blue, oxygen as red, phosphorus as magenta, cobalt as orange, and carbon as red (the cobalamin cofactor) or cyan (the UAG-encoded amino acid).
Table 4.1: Statistics for data collection, and phase determination. All the calculations were done with the program PHASES (103). For completeness and $R_{\text{sym}}$, numbers in parentheses represent the statistics for the shell comprising the outer 10% (theoretical) of the data. Phasing power is the mean value of the heavy atom structure factor divided by the lack of closure. Figure of merit is the mean value of the cosine of the error in phase angles. The combined figure of merit for all datasets was 0.63.
<table>
<thead>
<tr>
<th></th>
<th>Native 1</th>
<th>Native 2</th>
<th>NaI</th>
<th>I₂</th>
<th>EMTS</th>
<th>Me₂PbOAc</th>
<th>K₂OsC₄</th>
<th>CsCl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NaCl</td>
<td>(NH₄)₂SO₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data statistics**

- **Heavy atom**
  - concentration: 2 M, saturated, 1 mM, 5 mM, 2 mM, 4.5 M
  - soaking time: 20 min, 5 min, cocrystal, cocrystal, 24 hours, cocrystal
  - no. of sites: 27, 9, 6, 4, 3, 12
  - Wavelength (Å): 1.0000, 1.0000, 1.3869, 1.3869, 0.9924, 0.9488, 1.1402, 1.5497
  - Resolution (Å): 1.55, 1.7, 2.3, 3.4, 2.7, 2.9, 2.9, 3.2
  - Observations: 549727, 313202, 187164, 66528, 152898, 157617, 126772, 66423
  - Unique: 140499, 105886, 43796, 14232, 26712, 22397, 22544, 17063
  - Completeness (%): 97.3(93.7), 96.7(94.3), 97.4(80.2), 98.6(91.5), 98.6(91.5), 97.3(97.2), 98.6(90.7), 99.6(99.8)
  - Redundancy: 3.6, 4.6, 4.2, 4.6, 5.6, 7.2, 5.5, 4.0
  - Mosulicity (°): 0.42, 0.33, 0.42, 0.31, 0.48, 0.46, 0.17, 0.54
  - Rskin (°): 7.1(51.0), 4.9(28.1), 9.2(25.0), 15.0(29.0), 9.8(26.9), 11.5(28.1), 11.5(23.6), 14.6(33.6)

Table 4.1 (continued)
<table>
<thead>
<tr>
<th>Phasing statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20.26</td>
</tr>
<tr>
<td>Phasing power</td>
<td>1.81/1.30</td>
</tr>
<tr>
<td></td>
<td>(iso/eno)</td>
</tr>
<tr>
<td>R_{	ext{pM}} (centric)</td>
<td>0.53</td>
</tr>
<tr>
<td>R_{	ext{pM}} (iso/eno)</td>
<td>0.11/0.12</td>
</tr>
<tr>
<td>figure of merit</td>
<td>0.34</td>
</tr>
</tbody>
</table>

* $R_{	ext{pM}} (h) = \frac{\sum|I| - |I|}{\sum|I|}$, where $I$ is the mean intensity of the $i$ observations of reflection $h$.

† $R_{\text{pM}} = \frac{\sum|F_{\text{PH}} - F_{\text{P}}|}{\sum|F_{\text{PH}}|}$, where $F_{\text{PH}}$ is the structure factor of the derivative and $F_{\text{P}}$ is that of the native data. The summation is valid only for centric reflections.

‡ $\Delta_{\text{iso}} = \sum_{\text{iso}} \Delta_{\text{iso}}$ for isomorphous differences and $\sum_{\text{ano}} \Delta_{\text{ano}}$ for anomalous differences, where $\Delta_{\text{iso}}$ and $\Delta_{\text{ano}}$ are the isomorphous and anomalous lack of closure respectively, and $\Delta_{\text{iso}}$ is the isomorphous difference and $\Delta_{\text{ano}}$ is the Bijvoet difference. The summation is taken over all acentric reflections.

Table 4.1 Continued.
<table>
<thead>
<tr>
<th>Data set</th>
<th>Native 1, NaCl</th>
<th>Native 2, (NH₄)₂SO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 – 1.55</td>
<td>20.0 – 1.70</td>
</tr>
<tr>
<td>Total number of nonhydrogen atoms</td>
<td>3490</td>
<td>3490</td>
</tr>
<tr>
<td>Number of water molecules</td>
<td>464</td>
<td>533</td>
</tr>
<tr>
<td>R&lt;sub&gt;crystal&lt;/sub&gt;* (R&lt;sub&gt;free&lt;/sub&gt;)† (%)</td>
<td>17.6 (18.8)</td>
<td>16.0 (17.4)</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>21.4</td>
<td>15.1</td>
</tr>
<tr>
<td>Water molecules</td>
<td>36.5</td>
<td>34.2</td>
</tr>
<tr>
<td>rms bond length deviation (Å)</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>rms bond angle deviation (°)</td>
<td>1.198</td>
<td>1.176</td>
</tr>
</tbody>
</table>

* R<sub>crystal</sub> = 100 x Σ(|F<sub>obs</sub> - F<sub>calc</sub>)/Σ|F<sub>obs</sub>|, where F<sub>obs</sub> and F<sub>calc</sub> are the observed and calculated structure factors, respectively.
† R<sub>free</sub> is the same as R<sub>crystal</sub>, was calculated using 10% of the data excluded from refinement.

Table 4.2: Refinement statistics.
CHAPTER 5

Purification, crystallization, and structure determination of the nickel-containing carbon monoxide dehydrogenase (NiCODH) component of the acetyl coenzyme-A decarbonylase synthase (ACDS) complex from Methanosarcina barkeri

5.1 Introduction

Biological methanogenesis plays a significant role in carbon monoxide fixation and global carbon cycle (175, 176). About two thirds of the methane in nature is produced as a byproduct of the growth of methanogens on acetate (177, 178). This process occurs in nearly every conceivable anaerobic environment, including the rumen, the lower intestinal tract of humans, sewage digestors, landfills, and marine and freshwater sediments. Thus, elucidation of the mechanism involved in methanogenesis is fundamental to our understanding of the global carbon cycle.

One methanogenic species capable of acetogenic growth is Methanosarcina barkeri. Growth on acetate by this organism first involves its conversion to acetyl coenzyme-A (acetyl-CoA) (177, 179). Acetyl-CoA then serves as the substrate for the acetyl-CoA decarbonylase synthase (ACDS) complex (180). This ACDS complex catalyzes two fundamental reactions: (1) the cleavage of C-C bound in acetate to a methyl group and CO, and the transfer of the methylnyl group to a corrinoid protein and then
tetrahydrosarcinapterin (reaction 1), and (2) the reversible oxidation of this CO to CO2 (reaction 2). The net reaction for this process is shown in reaction 3.

\[
\text{acetyl-CoA} + \text{H}_4\text{SPt} \leftrightarrow \text{CH}_3\text{-H}_4\text{SPt} + \text{CoA} + \text{CO} \quad (1)
\]
\[
\text{CO} + \text{H}_2\text{O} \leftrightarrow \text{CO}_2 + 2\text{H}^+ + 2\text{e}^- \quad (2)
\]
\[
\text{acetyl-CoA} + \text{H}_4\text{SPt} + \text{H}_2\text{O} \leftrightarrow \text{CO}_2 + 2\text{H}^+ + 2\text{e}^- + \text{CH}_3\text{-H}_4\text{SPt} + \text{CoA} \quad (3)
\]

$\text{H}_4\text{SPt}$ and $\text{CH}_3\text{-H}_4\text{SPt}$ stand for tetrahydrosarcinapterin and $N^{\delta}$-methyltetrahydrosarcinapterin. The ACDS complex from $M. \text{barkeri}$ is comprised of five distinct subunits arranged into a $\alpha_8\beta_8\gamma_8\delta_8\varepsilon_8$ structure with a total molecular weight around 2000 kDa ($181, 182$). The individual sizes of each subunit are: $\alpha$ subunit, 89 kDa; $\beta$ subunit, 60 kDa; $\gamma$ subunit, 50 kDa; $\delta$ subunit, 48 kDa; $\varepsilon$ subunit, 20 kDa. This multimeric complex can be dissociated into three specific components during purification, each of which exhibits one of the fundamental activities of the whole complex ($181, 183, 184$). The reversible CO dehydrogenase (NiCODH) activity (reaction 2) has been localized at $\alpha$ and $\varepsilon$ subunits. A stable homodimeric $\alpha_2\varepsilon_2$ component with a molecular weight of 215 kDa can be isolated and purified. The reversible acetyl transfer and cleavage activity (reaction 1) is localized in the $\beta$ component. The reversible methylcobamide:tetrahydopterin methyltransferase activity (reaction 1) is localized in the $\gamma$ and $\delta$ subunits, which consist of a corrinoid protein and an iron-sulfur containing protein. The ACDS homologues are widely distributed in many anaerobic microorganisms, both bacteria and archaea. One of the best-known examples is the acetyl-CoA synthase from acetogens catalyzing the generation of acetyl-CoA via the
energy-yielding Ljungdahl-Wood pathway, which is in reverse analogy to the ACDS complex (183).

In addition to its important functional role in methanogenesis, the ACDS complex has drawn extra attention because it constitutes two of six known nickel-containing cofactors in nature, and because its reaction mechanism involves the formation of an organometallic bond (Ni-CH\textsubscript{3} bond in reaction 1), a theme for which there is only one other example in biology (Co-CH\textsubscript{3} bond in vitamin B\textsubscript{12}) (185, 186). For over two decades, numerous spectroscopic experiments have been carried out to identify the structures of these metal cofactors and to elucidate how the cofactors might mediate these reactions. So far, most studies were performed using the ACDS or NiCODH proteins from acetogens (\textit{Clostridium thermoaceticum}) and phototropic bacteria (\textit{Rhodosprillum rubrum}) (3, 183). Although the overall homology among the proteins from different microorganisms is not high, the cofactors that drive these reversible reactions are assumed to be similar. It has been widely believed that the two unique nickel centers in ACDS complex from methanogens are in a form of Ni-Fe-S clusters, and located in the \(\alpha\) (called A-cluster) and the \(\varepsilon\) (called C-cluster) subunits, respectively (183). The Ni ions in both clusters serve as either acetyl group (in A-cluster) or CO (in C-cluster) binding sites. A molecular tunnel through the enzyme connects these clusters (187, 188).

In an effort to understand the mechanisms of acetyl-CoA degradation and CO/CO\textsubscript{2} oxidation in methanogens, and to elucidate the structures of the Ni-Fe-S clusters, we have chosen to start with the structure determination of the \(\alpha_2\varepsilon_2\) NiCODH core of the \textit{M. barkeri} ACDS complex. There are three reasons for working with the structure of this component. First, previous studies have shown that this component can form a stable
dimer during the purification that exhibit the independent CO/CO₂ reversion activity. Secondly, this structure could be the first CODH structure from methanogens. Third, comparing to α subunit, very little is known about the properties and functions of the ε subunit. The structure of the α₂ε₂ dimer will provide useful information. Due to the extreme oxygen sensitivity and the large size of this protein, the purification, crystallization, and structure determination have required extensive effort. So far, two crystal forms of this protein have been obtained and two sets of MAD data have been collected with the resolution up to 2.2 Å. A total of nine metal clusters have been identified for the α₂ε₂ dimer through the Patterson search. While this project is in progress, two NiCODH structures from hydrogenogenic and phototropic bacterium have been solved (1, 2). While the location and arrangement of some of the clusters within these structures appear to exhibit similarity to our current results for the α₂ε₂ NiCODH core, there are also some differences as well.

5.2 Materials and methods

5.2.1 Preparation of cell extracts and purification of the α₂ε₂ NiCODH component

*Methanosarcina barkeri* strain MS (DSM 800) was cultured in 40 L carboys under anaerobic conditions in a phosphate-buffered medium supplemented with 80 mM sodium acetate (189). Cells were harvested by anaerobic centrifugation 14 days after inoculation, and washed three times with 50 mM MOPS buffer, pH 7.0, and 1 mM sodium dithionite, and stored at -70 °C until use. These cells were provided by Dr. Krzycki’s laboratory.
To prepare cell extracts, the frozen cells were first doused with liquid nitrogen and then blended for two minutes. After degassing for two hours, cells were suspended in MOPS buffer (1.25 gram wet weight per ml of buffer), followed by disruption with a French pressure cell prior to ultracentrifugation at 150,000 g for 2 hours. The supernatant was stored at –70 °C until use.

CODH was purified according to previously published procedures with only slight modifications (189). All the purification steps were performed in an anaerobic chamber containing 97% N₂ and 3% H₂ (Coy Laboratories Products Inc., Grass Lake, MI) at room temperature. Trace oxygen was eliminated by circulating the chamber atmosphere through palladium catalysts. All the buffers and column materials were made anaerobic by repeated cycles of evacuation and flushing with N₂. Sodium dithionite (1 mM) and dithiothreitol (1 mM) were added to the buffers prior to use. Cell extracts were loaded on to a DE-52 column (5.0 x 35 cm) equilibrated with 100 mM NaCl in 50 mM Tris-HCl buffer, pH 8.0. The column was eluted with a 2800 mL linear gradient of 100~500 mM NaCl at a flow rate of 3 mL/min. Fractions containing the most CODH activities were pooled and applied to a DEAE-Sepharose column (5.0 x 10 cm) that had been equilibrated in advance with 100 mM NaCl in 50 mM MOPS buffer, pH 7.0. The column was eluted with a 1500 mL linear gradient of 100~500 mM NaCl at a flow rate of 3 mL/min. The active fractions were combined and concentrated by Amicon and then loaded onto a Sepharose CL-6B column (2.6 x 85 cm) that had been previously equilibrated with 50 mM MOPS buffer, pH 7.0. The active fractions were then pooled and applied to a Mono-Q HR 10/10 column equilibrated with 100 mM NaCl in 50 mM MOPS buffer, pH 7.5. Elution was performed using a 100 mL linear gradient of 100~500
mM NaCl at a flow rate of 1ml/min. The final active pure protein was concentrated and stored at -80 °C. Protein concentrations were determined by the Lowry method using bovine serum albumin as standard (I90). The purity of the protein was checked by SDS-PAGE and dynamic light scattering experiments (Figure 5.1).

5.2.2 Enzyme assays

The NiCODH activity was measured based on the reduction of methyl viologen from CO oxidation monitored at 578 nM using a Beckman DU-70 spectrophotometer. All reactions were performed under N$_2$ (without H$_2$) as previously described (I89, I91). Briefly, each disposable cuvette was capped with a rubber stopper in the anaerobic chamber and then flushed and evacuated with N$_2$. The reaction mixture contains 50 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$ buffer (pH 7.0) and 5 mM methyl viologen (Sigma). The mixture (1 mL) and 0.4 mL of 100% CO was then added to the cuvette sequentially. After equilibration for about 15 min at room temperature, sodium dithionite solution was added to the cuvette. It turns light blue in the absence of oxygen. The reaction was initiated by adding the enzyme. The reduction rate was monitored at 578 nM for 30 seconds.

5.2.3 Crystallization and data collection

The crystallization of the $\alpha_2\epsilon_2$ NiCODH core was carried out in a Coy anaerobic chamber by using a modified version of batch capillary methods. For each condition, 15 µl of precipitating agents (supplemented with 1 mM dithionite) was injected gently on the top of the protein sample (100 mg/ml, 50 mM MOPS, pH 7.5, 200-300 mM NaCl, 10 µl) in the capillary without disturbing the interface. An alternative method is to place silica
hydrogel (Hampton Research) at middle of the capillary, and then add the protein and precipitating agents at each side of the gel. The capillaries were then sealed by epoxy and stored in the anaerobic chamber at room temperature.

Two crystals were used in structure analysis {form 1: 0.2 M (NH$_4$)$_2$SO$_4$, 25% PEG 4K, and 0.1 M Na acetate (pH 4.6); and form 2: 0.2 M Mg acetate, 20% PEG 1K, and 0.1 M Na cacodylate (pH 6.5)} (Figure 5.2). Prior to data collection, these crystals were frozen in liquid propane immediately without adding any cryoprotectant after opening the capillaries.

The native Fe MAD data sets from both crystal forms were collected on beamline X4A at the National Synchrotron Light Source in Brookhaven National Lab with the resolution of crystal form 1 up to 2.2 Å (Figure 5.3) (Table 5.1). The Xe and Kr derivative crystals (form 1) were generated by using a MRC Cryo-Xe-Siter at the beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL). The pre-frozen NiCODH crystals were thawed in paratone-N oil (Hampton Research), and then mounted on a cryo-loop and pressurized in the Cryo-Xe-Siter under a Xe or Kr pressure of 300 psi for 40 minutes. The Xe- and Kr-NiCODH crystals were flash-cooled under pressure in liquid carbon tetrafluoride. The Xe- (single wavelength) and Kr-NiCODH (multiwavelength) data sets were collected at the beamline 9-2 at SSRL (Table 5.2)

Crystal form 1 contains a αε monomer within the asymmetric unit and belongs to space group P2$_1$2$_1$2$_1$ with unit cell dimensions a = 81.7 Å, b = 101.2 Å, c = 133.2 Å. Crystal form 2 appears to contain the entire α$_2$ε$_2$ dimer within the asymmetric unit with unit cell dimensions a = 77.3 Å, b = 118.5 Å, c = 187.2 Å, $\alpha = \beta = \gamma = 90.0$. However, the space group of crystal form 2 remains uncertain due to the obscuration of the systematic
absence and thus could be any one of $P_{21}2_1$, $P_{21}2_12$, $P_{21}2_12_1$ or $P_{2}2_{1}2_{1}$. The Matthew’s coefficients for crystal form 1 and 2 are $2.59 \text{ Å}^3/\text{dalton}$ (for the $\alpha\varepsilon$ monomer) and $2.02 \text{ Å}^3/\text{dalton}$ (for the $\alpha_2\varepsilon_2$ dimer), respectively.

5.2.4 The Patterson search and MAD phasing

The isomorphous and anomalous difference Patterson maps for the Fe MAD data sets were generated using PHASES (103) (Figure 5.4). A total of five FeS cluster sites were located manually for crystal form 1. The Patterson maps generated from different combination of the data sets, resolutions and sigma cut-off levels were used to confirm the Fe sites. The Xe and Kr sites were located by difference Fourier using the Fe MAD phase. And in reverse, using Xe or Kr phases alone could find the same Fe cluster sites.

The initial phase for crystal form 1 was calculated by PHASES (103) and MLPHARE (154) from CCP4 program suite (155). The positions of the individual Fe were optimized manually according to the electron density map. The heavy atom parameters and phases were further refined by several programs including DM (156), CNS (103), SOLVE (192) (www.solve.lanl.gov), and SHARP (193).

5.3 Results and discussions

5.3.1 Purification of the enzyme

After the four chromatographic steps, the homogeneity of the final protein samples was checked by SDS-PAGE and dynamic light scattering (DLS) experiments (Figure 5.1). The SDS-PAGE gel shows a complex consisting of both $\alpha$ and $\varepsilon$ subunits. The DLS results indicated a monodispersed solution with an average molecular mass
around 200 kDa, which matches the molecular weight of the $\alpha_2\varepsilon_2$ dimer. The enzymatic activity was measured after every purification step and only the fractions with high activity were pooled and used for the following experiments.

5.3.2 The problems associated with crystallization and data collection

After several months to more than a year, dark brownish crystals were obtained from several different crystallization conditions (Figure 5.2). However, most crystals were useless due to high mosaicity, twinning, or an extremely long axis. So far, only two crystals were used in structure analysis. However, both crystals were very sensitive to the change of the environment. They tended to dissolve even when transferred to the drop made of the original mother liquid. Thus, these crystals were frozen by liquid propane without cryoprotectant.

After the initial screens, numeral trials have been set up to optimize the crystallization conditions in order to get more crystals. Basically, we have tried to vary several factors, including the pH, the type and concentration of salts, precipitating agents and additives, the concentration of the protein, the ratio between the protein and precipitating agent, and modification of crystallization methods. However, we were not able to find any more useful crystals under either similar or different conditions. This problem more than likely is rooted in the different quality of the proteins. I have purified four batches of proteins using presumably same procedures, but only the proteins from one of the batches can be crystallized. It seems that slight differences in the purity, the enzymatic activity, the oxidation state, and the aggregation state of the proteins affect the
crystallization dramatically. An efficient purification routine and sequentially immediate crystallization set-up should be helpful in getting new crystals in the future.

During data collection, several problems have also emerged. First, severe ice rings have been observed in some of the data sets from crystal form 1, especially the data around 3 to 4 Å, which is critical for identification of the individual Fe in the clusters (Figure 5.3). Secondly, a huge overlap of the diffraction spots in images has been observed in the data sets from crystal form 2 because of high mosaicity of the crystal and large oscillation angle. Third, in both crystal forms, crystals started to decay after a couple of data sets even under low temperature. Finally, due to technical reasons, the measurement of the Fe edge and peak for both crystal forms was not very accurate at the time. This could diminish the isomorphous and anomalous differences of the data sets. Since there are no more crystals available for either of the forms, all the refinements were limited to use the current data sets.

5.3.3 Identification of metal clusters and calculation of MAD phase

Five iron clusters within each $\alpha\varepsilon$ unit have been identified from the isomorphous and anomalous difference Patterson maps of crystal form 1 (Figure 5.4). These assignments are confirmed by the electron density calculated from the initial phases (Figure 5.5A). The distance between the centers of each adjacent cluster is around 11 to 15 Å, which suggests a possible pathway for electron transfer within the $\alpha\varepsilon$ unit (Figure 5.5B). One of the clusters is buried in the interface of the $\alpha_2\varepsilon_2$ dimer and shared by both monomers.
A bones trace has also been obtained from the initial electron density map that indicates a basic arrangement and shape for the $\alpha_2\varepsilon_2$ dimer core (Figure 5.6A). The core appears to be divided into two distinct $\alpha\varepsilon$ units, each containing four iron clusters with a joint cluster in the interface of the dimer. Unfortunately, the overall quality of the initial electron density does not appear to be good enough for positioning of the individual Fe atoms and for tracing the backbone. There are no secondary structural elements visible in the density maps.

The Patterson search of Fe sites in crystal form 2 was not successful. The molecular replacement searches also failed for crystal form 2, while using either the Fe sites or the electron density from crystal form 1 as the search model.

### 5.3.4 Comparison with other NiCODHs

As mentioned above, two crystal structures of NiCODH from *Carboxydothermus hydrogenoformans* and *R. rubrum* have been determined recently (1, 2). These two CODHs are both monofunctional enzymes and catalyze the same reaction as our $\alpha_2\varepsilon_2$ NiCODH core. The overall folds and the arrangement of the metal centers of these structures are virtually identical (Figure 5.6B). They both adopt a butterfly shape that is composed of two identical subunits. Each subunit consists of three domains, the N-terminal helical domain, the middle and the C-terminal $\alpha\beta$ Rossmann-like domains. Each dimer has five metal centers of three types: B, C, and D. Both B- and D-clusters are conventional cubane-type [4Fe-4S] clusters. The D-cluster is located at the surface of the dimer, with each subunit providing half of the [4Fe-4S] cluster. The B-cluster is sited
about 10 Å away from the D-cluster and deep down in the molecule. The most significant difference between these two structures is the structure of the C-clusters.

C-cluster is a mixed-type cluster that is presumably responsible for CO binding and oxidation. Previous spectroscopic studies have proposed that the C-cluster is composed of a Fe$_4$S$_4$ cubane cluster in which one of the Fe ions is bridged to an external Ni ion via an unknown ligand (“X”) (194) (Figure 5.7A). These two crystal structures of NiCODH, however, have offered two surprising models for the C-cluster (Figure 5.7B and C). In both models, the Ni ion is in fact an integral constituent of the FeS cubane, and instead a Fe ion is bridged to the cubane through a sulfur atom. The major difference between these two models, however, is the sulfur atom that bridges the Ni and the external Fe. In Dobbek’s model, the external Fe is bridged by a sulfide, while in Drennan’s model, it is the thiol group of a cysteine residue.

Although the overall homology between the $\alpha$ subunit of our protein and other two CODH is not high, the binding sites for metal clusters, including most of the protein ligands (Cys and His), are well conserved (Figure 5.8). Additionally, based on our preliminary results, the position and spatial arrangement of the metal clusters appear to exhibit some similarity. Five of the nine clusters in our $\alpha_2\varepsilon_2$ dimer can be superimposed with five clusters from two known structures (Figure 5.6). This resemblance is consistent with the data from spectroscopic and biochemical studies, and suggests a conserved electron transfer pathway for CO/CO$_2$ oxidation. On the other hand, there are some differences as well. First, some of the cysteine ligands of C- and D-clusters, which are boxed in Figure 5.8, are not conserved in our proteins, which could imply an alternative structure for the critical NiFeS and bridge clusters in methanogens. Secondly,
our structures appear to exhibit two extra sites for metal centers located at either side of the bridge cluster. The putative binding sites for these two clusters could be identified from the sequence analysis (Figure 5.7). Finally, the locations of the joint clusters (D-clusters) are different among CODHs. From our preliminary results, this cluster is buried in the interface of the two subunits. In two other structures, this cluster is located at the surface of the molecule (Figure 5.6). These differences may be due to the multiple functions of different CODH enzymes.

Attempts have been made to solve the structure of the $\alpha_2\epsilon_2$ NiCODH core by the molecular replacement method using the two known structures as search models. These efforts have so far been unsuccessful. The reasons could be the low overall homology, the larger size of our protein (214 kDa versus about 130 kDa for other two proteins), and the poor quality of the diffraction data. What will presumably be needed to solve this structure is a way to obtain more crystals reproducibly.
Figure 5.1: The quality of the $\alpha_2\varepsilon_2$ NiCODH component of the ACDS complex. (A) SDS-PAGE gel. From left to right: cell extract, $\alpha_2\varepsilon_2$ NiCODH, MW markers. (B) Dynamic light scattering measurement.
Figure 5.2: Crystals of the $\alpha_2\varepsilon_2$ NiCODH core of the *M. barkeri* ACDS complex.
Figure 5.3: Diffraction pattern of the crystal form 1 of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex.
Figure 5.4: Harker sections of the isomorphous difference Patterson maps of the α2ε2 NiCODH core of the ACDS complex. The peaks for the metal clusters are labeled by stars.
Figure 5.5: The identified metal centers of the $\alpha_2\varepsilon_2$ NiCODH core of the ACDS complex. (A) The electron density of the metal clusters. (B) The corresponding distances between the clusters. The clusters from two different subunits are colored by green-yellow and purple-yellow, respectively.
Figure 5.6: Comparison of the $\alpha_2\epsilon_2$ NiCODH core of the ACDS complex with two known CODH structures. (A) The molecular surface of the $\alpha_2\epsilon_2$ core. The bones traces of two subunits are colored cyan and brown, respectively. The clusters are colored as Figure 5.5. (B) Superimposed two known CODH structures. PDB ID: 1JJY (pink, and red), and 1jqk (cyan and blue).
Figure 5.7: The structures of C-clusters. (A) Proposed structure of C-cluster; (B) Dobbek’s model. (C) Drennan’s model.
Figure 5.8: Sequence alignment of the protein regions interacting with the metal clusters in NiCODH family. The conserved residues are colored according to the interaction with the clusters: cyan, cluster D; green, cluster B; purple, cluster C; orange, the extra clusters in CODH from mthanogens; blue, the conserved residues adjacent to the clusters and may involve in the electron transfer; boxed, the missing residue in *M. barkeri*. The sequence of the *M. barkeri* CODH was determined by Dr TC Tallant from Dr. Krzycki’s lab. The residue numbers of *M. barkeri* CODH are labeled above the sequences. The other sequences were obtained from the SWISS-PROT/TrEMBL database ([http://us.expasy.org/sprot/sprot-top.html](http://us.expasy.org/sprot/sprot-top.html)) or adopted from the papers (1-3).
Figure 5.8
Table 5.1: Summary of data collected for crystal form 1 and 2 of the \( \alpha_2\varepsilon_2 \) NiCODH core of the ACDS complex.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>form 1</th>
<th>form 1</th>
<th>form 1</th>
<th>form 1</th>
<th>form 1</th>
<th>form 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.5500</td>
<td>1.6000</td>
<td>1.7418</td>
<td>1.7415</td>
<td>1.7360</td>
<td>1.7360</td>
</tr>
<tr>
<td>anomalous</td>
<td>Fe-remote1</td>
<td>Fe-remote2</td>
<td>Fe-edge1</td>
<td>Fe-edge2</td>
<td>Fe-peak1</td>
<td>Fe-peak2</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>101.157</td>
<td>101.527</td>
<td>100.773</td>
<td>100.945</td>
<td>101.305</td>
<td>101.185</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 - 2.2</td>
<td>20.0 - 2.5</td>
<td>20.0 - 2.5</td>
<td>20.0 - 2.5</td>
<td>20.0 - 2.8</td>
<td>20.0 - 2.6</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>213525</td>
<td>187159</td>
<td>143854</td>
<td>160579</td>
<td>99069</td>
<td>172281</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>53308</td>
<td>36642</td>
<td>36920</td>
<td>37200</td>
<td>26957</td>
<td>32538</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>93.7 (97.3)</td>
<td>93.2 (99.7)</td>
<td>96.1 (99.8)</td>
<td>96.5 (100.0)</td>
<td>97.3 (99.7)</td>
<td>95.1 (100.0)</td>
</tr>
<tr>
<td>Rmerge (%)†</td>
<td>11.2 (28.5)</td>
<td>11.6 (32.6)</td>
<td>9.30 (18.4)</td>
<td>9.30 (21.0)</td>
<td>12.4 (26.1)</td>
<td>13.7 (27.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crystal</th>
<th>form 2</th>
<th>form 2</th>
<th>form 2</th>
<th>form 2</th>
<th>form 2</th>
<th>form 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P212121</td>
<td>P212121</td>
<td>P212121</td>
<td>P212121</td>
<td>P212121</td>
<td>P212121</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.7367</td>
<td>1.7428</td>
<td>1.7068</td>
<td>1.7428</td>
<td>1.7353</td>
<td>1.7432</td>
</tr>
<tr>
<td>anomalous</td>
<td>Fe-peak1</td>
<td>Fe-edge1</td>
<td>Fe-remote1</td>
<td>Fe-edge2</td>
<td>Fe-peak2</td>
<td>Fe-edge3</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>118.461</td>
<td>118.091</td>
<td>119.362</td>
<td>120.136</td>
<td>121.045</td>
<td>120.804</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 - 3.0</td>
<td>20.0 - 3.1</td>
<td>20.0 - 3.0</td>
<td>20.0 - 2.8</td>
<td>20.0 - 3.5</td>
<td>20.0 - 3.3</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>79459</td>
<td>76383</td>
<td>78813</td>
<td>81712</td>
<td>53197</td>
<td>67746</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>30775</td>
<td>30629</td>
<td>30980</td>
<td>41175</td>
<td>20771</td>
<td>26942</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>90.3 (79.8)</td>
<td>91.4 (83.3)</td>
<td>91.1 (80.7)</td>
<td>92.4 (87.2)</td>
<td>81.0 (73.8)</td>
<td>96.5 (95.1)</td>
</tr>
<tr>
<td>Rmerge (%)†</td>
<td>14.7 (27.0)</td>
<td>13.0 (28.8)</td>
<td>11.3 (27.9)</td>
<td>7.8 (28.4)</td>
<td>11.5 (26.4)</td>
<td>7.8 (27.3)</td>
</tr>
</tbody>
</table>

* The numbers in parentheses for the highest resolution shell.
† \( R_{sym}(I) = \Sigma_h \Sigma_i |I_i-\bar{I}|/\Sigma_h \Sigma_i I, \) where \( I \) is the mean intensity of the \( i \) observations of reflection \( h \).
<table>
<thead>
<tr>
<th>Crystal</th>
<th>codkr1</th>
<th>codkr2</th>
<th>codkr3</th>
<th>codkr4</th>
<th>codkr5</th>
<th>codkr6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.8660</td>
<td>0.8265</td>
<td>0.8652</td>
<td>0.8655</td>
<td>0.8321</td>
<td>1.7389</td>
</tr>
<tr>
<td>anomalous</td>
<td>Kr-edge1</td>
<td>Kr-remote1</td>
<td>Kr-peak1</td>
<td>Kr-edge2</td>
<td>Kr-remote2</td>
<td>Fe-peak2</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>a = 82.411</td>
<td>a = 82.419</td>
<td>a = 82.429</td>
<td>a = 82.392</td>
<td>a = 82.559</td>
<td>a = 82.484</td>
</tr>
<tr>
<td></td>
<td>b=101.364</td>
<td>b=101.392</td>
<td>b=101.079</td>
<td>b=101.111</td>
<td>b=101.124</td>
<td>b=101.207</td>
</tr>
<tr>
<td></td>
<td>c = 133.419</td>
<td>c = 133.442</td>
<td>c = 133.154</td>
<td>c = 133.171</td>
<td>c = 133.253</td>
<td>c = 133.350</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 – 2.3</td>
<td>20.0 – 2.3</td>
<td>20.0 – 3.0</td>
<td>20.0 – 2.8</td>
<td>20.0 – 3.0</td>
<td>20.0 – 2.8</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>160291</td>
<td>155945</td>
<td>79027</td>
<td>82750</td>
<td>77666</td>
<td>89091</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>48730</td>
<td>47679</td>
<td>23155</td>
<td>26882</td>
<td>28121</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>96.8 (92.3)</td>
<td>94.7 (86.8)</td>
<td>99.4 (99.9)</td>
<td>94.2 (95.9)</td>
<td>99.1 (99.8)</td>
<td>99.4 (99.6)</td>
</tr>
<tr>
<td>Rmerge (%)†</td>
<td>7.7 (22.6)</td>
<td>9.9 (32.1)</td>
<td>13.9 (30.0)</td>
<td>12.5 (30.9)</td>
<td>14.1 (31.3)</td>
<td>11.2 (33.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crystal</th>
<th>codxe</th>
<th>codxe2</th>
<th>codxe3</th>
<th>codxe4</th>
<th>codxe5</th>
<th>codxe6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
<td>P21221</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.907</td>
<td>1.698</td>
<td>1.742</td>
<td>1.907</td>
<td>1.5497</td>
<td>1.736</td>
</tr>
<tr>
<td>anomalous</td>
<td>Xe-peak</td>
<td>Fe-remote1</td>
<td>Fe-edge</td>
<td>Xe-peak2</td>
<td>Fe-remote2</td>
<td>Fe-peak</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>a = 82.235</td>
<td>a = 82.216</td>
<td>a = 82.318</td>
<td>a = 82.408</td>
<td>a = 83.106</td>
<td>a = 83.261</td>
</tr>
<tr>
<td></td>
<td>b=101.289</td>
<td>b=101.324</td>
<td>b=101.357</td>
<td>b=101.410</td>
<td>b=101.630</td>
<td>b=101.732</td>
</tr>
<tr>
<td></td>
<td>c = 133.367</td>
<td>c = 133.448</td>
<td>c = 133.708</td>
<td>c = 133.863</td>
<td>c = 133.629</td>
<td>c = 134.048</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 – 3.0</td>
<td>20.0 – 2.7</td>
<td>20.0 – 2.7</td>
<td>20.0 – 3.0</td>
<td>20.0 – 2.5</td>
<td>20.0 – 2.9</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>64967</td>
<td>85254</td>
<td>94261</td>
<td>62968</td>
<td>129204</td>
<td>71665</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>21519</td>
<td>29460</td>
<td>30642</td>
<td>22172</td>
<td>39561</td>
<td>25537</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>94.4 (96.4)</td>
<td>94.2 (77.8)</td>
<td>97.7 (96.2)</td>
<td>96.1 (99.2)</td>
<td>99.3 (99.2)</td>
<td>99.3 (99.2)</td>
</tr>
<tr>
<td>Rmerge (%)†</td>
<td>17.8 (28.3)</td>
<td>8.8 (19.9)</td>
<td>10.5 (31.6)</td>
<td>11.5 (30.3)</td>
<td>7.0 (28.5)</td>
<td>14.1 (30.5)</td>
</tr>
</tbody>
</table>

* The numbers in parentheses for the highest resolution shell.
† Rsym (I) = Σ_h [Σ_i |I_i| - |I|] / Σ_h |I|, where I is the mean intensity of the i observations of reflection h.

Table 5.2: Data processing statistics for Kr and Xe derivatives of the \( \alpha_2\varepsilon_2 \) NiCODH core of the ACDS complex.


166


