# Structural Insight into the Assembly of Iron-Sulfur Clusters and their Function in Radical Generation

by

Jessica L. Vey

B.S., Chemistry (2001)

Temple University

Submitted to the Department of Chemistry in Partial
Fulfillment of the Requirements for the Degree of Doctor
of Philosophy in Biological Chemistry

at the

Massachusetts Institute of Technology

February 2008

Signature of Author _____

<div align="right">Department of Chemistry<br>October 31<sup>st</sup>, 2007</div>

Certified By _____

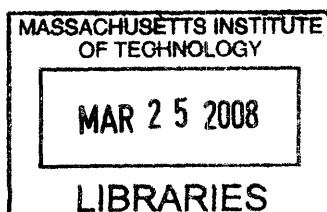<div align="right">Catherine L. Drennan<br>Associate Professor of Chemistry<br>Thesis Supervisor</div>

Accepted By _____

<div align="right">Robert W. Field<br>Chairman, Department Committee on Graduate Students</div>

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:


Professor JoAnne Stubbe_____            _____
                                                                          Committee Chairman


Professor Catherine L. Drennan_____          ___
                                                                          Thesis Supervisor


Professor Stuart Licht_____
                                                                          Committee Member

# Structural Insight into the Assembly of Iron-Sulfur Clusters and their Function in Radical Generation

by

Jessica L. Vey

## ABSTRACT

This thesis addresses two emerging areas in the study of iron-sulfur cluster biochemistry: bioassembly of iron-sulfur clusters, and their involvement in initiation of radical chemistry. The structure of a cysteine desulfurase involved in cluster bioassembly in the cyanobacterium *Synechocystis* PCC sp. 6803 was solved by X-ray crystallography and analyzed in terms of its mechanistic implications. We found that the active site cysteine responsible for the direct removal of sulfur from substrate cysteine is located on a short, well-ordered loop, consistent with structures solved of homologous proteins. The length of this loop is thought to restrain the active site cysteine, interfering with its ability to affect catalysis. Our results are consistent with the theory that this cysteine desulfurase requires an accessory protein for fully activity *in vivo*.

Two structures of pyruvate formate-lyase activating enzyme from *Escherichia coli*, an *S*-adenosylmethionine radical enzyme, were also solved by X-ray crystallography, providing the first structure of an activase from this family of enzymes. These structures revealed the enzyme's active site and the residues involved in binding and orienting substrate for hydrogen atom abstraction. Comparison of the structures of the substrate-free and substrate-bound forms of the enzyme identified a conformational change associated with substrate binding. Detailed analyses of the structure of pyruvate formate-lyase activating enzyme were carried out to provide insight into catalysis. These structures were also analyzed in comparison with other *S*-adenosylmethionine radical enzyme structures to more clearly understand the structural basis for reactivity in this superfamily.

Thesis Supervisor: Catherine L. Drennan
Title: Associate Professor of Chemistry

# ACKNOWLEDGEMENTS

TABLE OF CONTENTS <span style="float:right">page</span>

1. AdoMet conformation
2. General properties of the AdoMet binding site
3. Overall description of the AdoMet binding site
      a. The methionyl moiety
      b. The ribose
      c. The adenine moiety
      d. AdoMet binding motifs
4. Deviations and variations in AdoMet binding
5. Implications of the AdoMet binding site architecture
6. Other known AdoMet-binding protein folds

1. Overall description of substrate binding site
2. The substrate binding sites
      a. PflAE
      b. HemN
      c. BioB
      d. MoaA
      e. LAM
3. Conformational changes associated with substrate binding
4. Similarities and differences in the substrate binding sites

1. Further studies of iron-sulfur cluster bioassembly.
2a. Detailed account of the PflAE crystallography.

7

**Overview of this Thesis**

Since the discovery of iron-sulfur clusters in the 1960s these cofactors have proven to be very common, versatile and important biological tools. With wide-ranging capabilities and many different forms, study of these uniquely useful cofactors has been, and continues to be, an exciting branch of biochemistry that attracts scientists from across multiple disciplines. In this thesis, we address two emerging areas in the study of iron-sulfur cluster biochemistry: bioassembly of the clusters themselves, and their newly-discovered ability to initiate radical chemistry. In Chapter I, we introduce these two topics and attempt to put them into a broader context. First, we briefly summarize our current state of knowledge on the iron-sulfur cluster synthetic process, focusing on the specific pathway relevant to the work presented herein. Second, we switch gears slightly to discuss the $S$-adenosylmethionine (AdoMet) radical enzyme superfamily and its use of an iron-sulfur cluster to initiate radical chemistry.

Chapter II presents the structure of SufS from the cyanobacterium *Synechocystis* PCC sp. 6803, cysteine desulfurase involved in cluster bioassembly in that organism. The structure is then discussed in terms of its mechanistic implications. Our cysteine desulfurase structure is similar to available structures of other SufS cysteine desulfurases, the key finding being that the active site cysteine, $C_{372}$, resides on a constrained loop in the active site. It is thought that the activity of the SufS desulfurases is impaired by the inflexibility of this loop, which leads to reliance on accessory factors to enhance activity. Serendipitously, a detergent molecule added to our crystallization conditions was observed bound at an interesting location in the SufS structure, suggesting a potential site for interaction with an accessory factor.

Chapters III and IV focus on catalysis by the AdoMet radical enzymes. In Chapter III we present two structures of pyruvate formate-lyase activating enzyme (PflAE), describe their implications on catalysis and discuss a theoretical docking model between PflAE and its substrate enzyme. We have solved structures of the substrate-free and substrate-bound forms of PflAE, and as a result have identified a conformational change associated with substrate binding. This conformational change is likely conserved in all of the AdoMet radical activating enzymes. These structures also provide a first view of the active site of this enzyme, and identify the specific residues involved in substrate binding and orientation. Finally, based on the orientation of the substrate in our structure, we generated a docking model between our PflAE structure and a fragment of the previously solved model of pyruvate formate-lyase (Pfl). This docking model may reflect the mode of interaction between PflAE and its full substrate, and provides a starting place for further experimentation focused on clarification of the activation reaction and any conformational changes that the reaction involves.

Chapter IV focuses on the five AdoMet radical structures currently available: PflAE, coproporphyrinogen III oxidase, biotin synthase, MoaA, and lysine 2,3-aminomutase. By comparing and contrasting each in terms of overall fold and active site, we hope to provide insight into the conserved aspects of catalysis by the AdoMet radical superfamily and the main differences between the individual enzymes that enable such diverse overall reactions. This chapter was written for publication as a review article.

The appendices in Chapter V describe additional experiments carried out related to the iron-sulfur cluster biosynthetic enzyme studied in Chapter II (Appendix 1) and provide details about the PflAE structure determination (Appendix 2). In Appendix 1, attempts to crystallize the SufS/SufE complex are described. Proteins from multiple

organisms (*Escherichia coli* and *Synechocystis*) were used to attempt to obtain a closely associated complex of SufS/SufE. Subsequently, several cyanobacterial scaffold proteins were cloned, expressed and purified with the goal of structurally characterizing a scaffold protein and / or the complex between a scaffold protein and cysteine desulfurase.

Appendix 2 describes in detail the crystallographic challenges encountered over the course of the structure determination of PflAE. All datasets, attempts at phasing, and problems during refinement are described, along with the steps taken to overcome each obstacle. This appendix also contains complete data processing statistics for most of the datasets collected, the individual heavy atom sites with statistics for each phasing method tried, and an explanation of the rationale used to properly build the peptide substrate of PflAE.

## Abbreviations used in this Thesis

| | |
|---|---|
| 5,6LAM | Lysine 5,6-aminomutase |
| 5'-dA | 5'-deoxyadenosine |
| 5'-dA• | 5'-deoxyadenosyl radical |
| 5'-GTP | Guanosine 5'-triphosphate |
| ABC | ATP binding cassette |
| AdoMet | $S$-adenosylmethionine |
| AdoCbl | adenosylcobalamin |
| aRNR | Class III ribonucleotide reductase |
| BioB | Biotin synthase |
| Bss | Benzylsuccinate synthase |
| CD | Cysteine desulfurase |
| C-DES | Cystine lyase |
| CoA | Coenzyme A |
| DTB | Dethiobiotin |
| DTT | Dithiothreitol |
| ecSufS | Group II cysteine desulfurase from *Escherichia coli* |
| FeS | Iron-sulfur (cluster) |
| Gdh | Glycerol dehydratase |
| GRE | Glycyl radical-containing enzyme |
| HemN | Coproporphyrinogen III oxidase |
| Hpd | 4- hydroxyphenylacetate decarboxylase |
| LAM | Lysine 2,3-aminomutase |
| LipA | Lipoyl synthase |
| MAD | Multiwavelength anomalous dispersion |
| MoaA/MoaC | Two specific enzymes involved in molybdenum cofactor biosynthesis |
| MR | Molecular replacement |
| PDB | Protein Data Bank |
| Pfl | Pyruvate formate-lyase |
| PflAE | Pyruvate formate-lyase activating enzyme |
| PLP | Pyridoxal 5'-phosphate |
| RCSB | Research Collaboratory for Structural Bioinformatics |
| RNR | Ribonucleotide reductase |
| SPL | Spore photoproduct lyase |
| sySufS | Group II cysteine desulfurase from *Synechocystis* PCC sp 6803 |
| TIM | Triosephosphate isomerase |
| TYW1 | AdoMet radical enzyme involved in wyobutosine biosynthesis |
| wt | wild type |

# Chapter I: Introduction

## I.A. Overview

Iron-sulfur (FeS) clusters are widespread and essential protein cofactors observed in all three kingdoms of life[1-3]. Multiple independent processes, which include respiration, gene regulation, nitrogen fixation and many other metabolic functions[1-5], require one or more forms of these inorganic cofactors. Their widespread occurrence and involvement with fundamental metabolic processes suggest that they are ancient cofactors[3]. FeS clusters are able to perform a wide range of functions such as electron transfer, structural stabilization, substrate recognition, catalysis and radical generation[1]. The clusters themselves can assume a variety of structures, some relatively simple, such as the ubiquitous 2Fe-2S and 4Fe-4S clusters, and some quite complicated[6-8], as observed in enzymes like nitrogenase[9-11], hydrogenase[12,13] and carbon monoxide dehydrogenase / acetyl-coA synthase [14-16] (Figure I.1).

The structures, chemical properties and biochemistry of protein-bound FeS clusters have been fields of very active study since discovery of the clusters in the 1960s[1]. Two new topics in FeS cluster biochemistry that have come under focus in recent years include the bioassembly the clusters themselves[4,17-23], as well as their role in radical generation by the AdoMet radical enzymes[24-40]. Because both of these topics have been reviewed extensively in recent years[4,17-40], this chapter will not serve as a comprehensive review, but will instead present sufficient background to explain how the work in this thesis fits in with our current understanding of the two fields.

## I.B. FeS cluster bioassembly

With the discovery of proteins involved in the biosynthesis of the nitrogenase cluster by

Dean and coworkers in 1993[41], FeS cluster bioassembly has become a subject of great

interest to scientists across many disciplines. Perhaps because of this broad appeal, the

field has witnessed an explosion in the amount of ongoing multidisciplinary research

aimed at further elucidating the assembly process[20]. Since it has been the subject of

many extensive and articulate reviews[4,17-23], here we present a brief summary of our

current understanding of FeS cluster synthesis in biology in general, with special focus on

the Suf biosynthetic pathway.


### I.B.1. Overview of FeS cluster bioassembly

Though they are, at least in some cases, readily reconstituted *in vitro* into

apoproteins using aqueous iron(II) and sulfide, FeS cluster biosynthesis *in vivo* is a

tightly controlled process. Presumably this tight control is due to the toxicity of the high

levels of iron and sulfide ions required for spontaneous assembly, and perhaps also

because conversion of sulfide into its persulfidic forms allows much greater control and

ease in transfer by proteins[42]. Overall, the mechanism of cluster bioassembly appears to

be conserved from prokaryotes to higher eukaryotes[43,44], an observation that suggests

common ancestry, consistent with the theory that these cofactors are quite ancient. After

several years of active research, many participating proteins in several different gene

clusters have been identified and characterized.

The first enzyme in this process to be isolated was NifS from *Azotobacter*

*vinelandii*, a cysteine desulfurase involved in the bioassembly of the nitrogenase complex

iron sulfur cofactor[41]. Other open reading frames within the gene cluster (designated *nif*

for nitrogen fixation) were also found to be involved in the biosynthesis of the nitrogenase cofactor. Subsequent experiments identified two additional gene clusters (Figure I.2), designated *isc* (iron sulfur cluster assembly) and *suf* (for sulfur mobilization / utilization) that have some homology to the *nif* operon and are involved in FeS cluster assembly. Proteins encoded by the *isc* operon are responsible for the majority of FeS cluster biosynthesis in some organisms, mainly bacteria and in the mitochondria of eukaryotes. The Suf proteins, which have been identified in bacteria and in plant chloroplasts, can serve multiple functions, depending on the organism under study. In *Escherichia coli* and *Erwinia chrysanthemi*, the Suf proteins seem to be used to repair oxidatively damaged clusters[45]. In these organisms, they can also act as a backup system for cluster bioassembly under conditions of oxidative stress and/or iron limitation[46]. However, in cyanobacteria such as *Synechocystis* and in the chloroplasts of *Arabadopsis thaliana*, Suf serves as the main source of FeS clusters[46-49]. Over the course of evolution, the Isc and Suf systems seem to have been incorporated from primitive organisms into eukaryotic genomes via endosymbiosis, giving rise to the related systems we have observed in various eukaryotic organelles, specifically chloroplasts and mitochondria[21,46]. Obviously, the question of the evolution of the multiple systems used for FeS cluster bioassembly will be slowly clarified as more is learned about the individual pathways. In addition, as we collectively learn more about the different biosynthetic pathways, we hope to clarify the differences that exist between them that allow the observed functional specialization.

*I.B.2. The basic components of an FeS bioassembly system*

Several basic components of FeS cluster biosynthesis have been identified and characterized (Figure I.3), including sulfur donor, chaperone and scaffold proteins, and electron transfer proteins. Cysteine desulfurases (Nif/Isc/SufS) provide sulfur for incorporation into clusters. The desulfurases, along with an unknown iron donor, are thought to build a cluster into a scaffold protein (Nif/IscU, and possibly Isc/SufA), after which the cluster is transferred to target proteins. In addition, a ferrodoxin-like protein encoded in all three gene clusters may be used to provide electrons at any of several steps, including transfer of iron from the iron delivery protein(s), the actual cluster formation reaction or cluster release from the scaffold and transfer to the target protein[46]. Proteins homologous to the molecular chaperone proteins DnaK and DnaJ (denoted HscB and HscA in Figure I.2) are also involved in bioassembly by the Isc machinery, possibly through transfer of the cluster, stabilization of the scaffold protein or the scaffold protein/desulfurase complex, or another less obvious activity[4,50]. Though the *suf* operon does not encode these molecular chaperones, such proteins could be encoded elsewhere in the genome. Finally, several *suf* operons encode an ABC-type ATPase, which may provide energy for FeS cluster biosynthetic machinery. The nomenclature used above applies to bacterial systems and will be used throughout this chapter for clarity, but does not apply to all organisms. Because the work in this thesis focuses on the cysteine desulfurase of the Suf system, the following section will discuss these proteins more specifically, using the proteins from *E. coli* as a model system.

*I.B.3. The Suf machinery*

Each of the three sets of FeS cluster biosynthetic proteins seem to follow the general

scheme outlined in Figure I.2. In the Suf pathway, sulfur is provided by a cysteine

desulfurase (SufS), though the observed low activity of SufS in comparison to IscS (the

Isc cysteine desulfurase) *in vitro*[48,51,52] suggests the involvement of an additional

activating protein[48]. SufE, a homodimeric protein with limited structural similarity to the

Isc scaffold protein (IscU)[19], is proposed to fulfill this role, forming a complex with SufS

and stimulating its cysteine desulfurase activity up to 40-fold[53,54]. A conserved cysteine

of SufE is essential for this function, though the precise mechanism is unclear[51,53-55].

Stimulation of SufS by SufE may be another mechanism of regulation of cluster

bioassembly by the Suf machinery and may also be related to the ability of this system to

operate under oxidative conditions.

The next step of the cluster biosynthetic process is the assembly of the cluster on

a scaffold protein. Extensive experimentation has led to the hypothesis that SufA is the

scaffold protein of this pathway, though some have proposed (on the basis of experiments

done on *E. coli* IscA) that it is actually the iron donor for the system[56-58]. For several

reasons discussed in section I.B.4, the initial hypothesis (that SufA is a scaffold protein)

seems more likely.

The *suf* operon lacks the two Isc chaperone proteins HscA and HscB (though

homologues could be encoded elsewhere in the genome). The function of HscA and

HscB are currently unclear, but they are known to interact with IscU[59-61] and likely

participate in the transfer of the cluster to apoproteins. Instead, the *suf* operon encodes

three proteins of unknown function, SufB, C and D, which form a complex, SufBCD[55,62].

SufC has homology to ABC-type ATPases[47,62,63], but SufB and SufD do not contain any

transmembrane motifs and have been found in the cytoplasm[55,62], ruling out the

possibility that they form an ABC transporter. Very little is known about the functions of SufBCD, though some have proposed that it provides energy for the Suf system and / or aids in similar transfer steps as the Isc chaperone proteins[46].

The Suf system also seems to differ from Isc in terms of regulation. The *isc* operon is regulated by an FeS cluster-containing transcription factor, IscR, that represses expression of the Isc proteins when in its cluster-bound state[64]. In *E. coli*, where the *suf* operon operates in response to iron limitation and oxidative conditions, Suf is under the control of Fur (which senses iron limitation) and OxyR (which activates the oxidative stress response)[4,65,66]. Regulation of Suf also varies between different organisms. In cyanobacteria, a specific regulatory protein called SufR was identified that acts as a transcriptional repressor of the operon[67-69], similar to IscR in *E. coli*. The differences between specific organisms in their mechanism of regulation of the *suf* operon may reflect the different roles this operon can play, as the main cluster biosynthetic machinery in some organisms (i.e. cyanobacteria) and as the "backup" machinery used during oxidative or iron-limited conditions (as in *E. coli*).


*I.B.4. Questions arising from our current understanding of FeS cluster biosynthesis*
FeS cluster biosynthesis is an amazingly complicated process, and extensive research has defined some of its basic characteristics. Perhaps because of the importance of this particular type of cofactor, organisms typically encode redundant biosynthetic systems and even multiple copies of the individual components. Accordingly, complete definition of any specific system tends to be problematic. The functional relationships between the multiple pathways – Isc, Suf and Nif – form another unknown that research in the field is slowly beginning to clarify. Indeed, the specificity of the *nif* operon gene products for

the nitrogenase cluster suggests the existence of additional operons encoding proteins for the assembly of other specific complex clusters, such as the A-cluster of CODH or the H-cluster of hydrogenase, though no specific operons for these purposes have been identified to date.

The chemical details of cluster formation are also incompletely characterized at this point. First, controversy surrounds the identity of the protein responsible for providing iron to the scaffold protein[46,56-58,70]. The general consensus is that frataxin (or the frataxin homolog cyaY in *E. coli*) fills this role[70], though some have suggested IscA and/or SufA as a possible candidate in *E. coli*[56-58]. However, both aerobic and anaerobic purification of SufA yields protein in the iron-free form, which can then be reconsitituted to contain FeS clusters[53]. These clusters reconstituted into SufA can be readily transferred to other apoproteins[53,71], suggesting a more likely role for this protein as a scaffold for assembly. Further, not all IscA proteins even bind iron tightly[72,73], although this finding is still controversial[74]. The subsequent steps of the formation reaction are also only vaguely understood. Because FeS clusters can form spontaneously, *in vitro* experiments aimed at elucidating the mechanism of assembly on the scaffold protein are particularly difficult. A reasonable mechanism has been proposed for both the order of addition to the scaffold protein and the reduction steps involved, in which two persulfidic sulfurs are transferred from the desulfurases to cysteine residues of the scaffold protein, followed by chelation of two ferrous ions by the persulfides. Electrons provided by the ferrous ions and by a redox active thiol nearby could then enable cluster formation[46]. However, further experimentation is necessary to more completely characterize the mechanism.

Another area of active research is the structural characterization of a scaffold protein[75-81], which has proven difficult. Excitingly, the first successful crystallization of a scaffold protein containing an FeS cluster was recently reported[81]. Finally, more work is needed to clarify how the transfer of an FeS cluster to its target protein occurs and what functions the other proteins involved in this process have. The role of HscA and HscB in the FeS cluster assembly process is being pursued by the Vickery laboratory[50], who have shown that HscA interacts with a specific sequence of IscU[59-61] and have solved structures of both chaperones, including HscA bound to the IscU recognition peptide[82,83].

Although extensive work has been carried out in this field, FeS cluster biosynthesis has proven to be astonishingly complex. More than a few questions remain to be answered concerning the different FeS cluster biosynthetic pathways. The work presented in Chapter II concerns the cysteine desulfurase of the Suf system, and may have implications for the ability of Suf to operate under stress conditions. We have solved the structure of SufS from Synechocystis PCC sp. 6803, which is the only essential desulfurase from that organism. As in other structures of this enzyme from different organisms, the active site cysteine resides on a short, presumably constrained loop. Restriction of the mobility of the active site cysteine may be the main cause of the catalytic inefficiency observed for the SufS enzymes.

**I.C. Radical generation by the AdoMet radical enzymes**

A newly identified biochemical capability of FeS clusters is their ability to mediate radical formation. The subject of Chapter III of this thesis, pyruvate formate-lyase activating enzyme (PflAE), is a well known member of the relatively new class of

enzymes termed the AdoMet radical superfamily[84]. In this section, we will introduce the superfamily and discuss the shared characteristics of its members, describe the structure of the central domain of the AdoMet radical enzymes, and then briefly focus on PflAE, or more specifically, on the protein substrate of PflAE, pyruvate formate-lyase (Pfl), to put this work into a greater context.

### I.C.1. Characteristics of the AdoMet radical enzymes

Members of this superfamily share a common mechanism to affect the initiation of radical chemistry through generation of the highly reactive 5'-deoxyadenosyl radical (5'-dA•)[26,30,33,39]. Towards this end, AdoMet radical enzymes bind a 4Fe-4S cluster via three cysteine residues within a conserved motif, and AdoMet ligates the unique iron of the cluster through its carboxylate oxygen and amino nitrogen atoms (Figure I.4)[85-90]. Binding of AdoMet directly to the cluster in this way brings its sulfonium sulfur atom into close proximity with the cluster. One electron reduction of AdoMet by, in most cases, flavodoxin[91-94] is believed to proceed via inner sphere electron transfer through the 4Fe-4S cluster to the AdoMet sulfonium, resulting in cleavage of AdoMet to yield 5'-dA• and methionine[39,89,95].

At this point during catalysis, the mechanisms of the individual AdoMet radical enzymes diverge to accomplish a wide variety of reactions on a similarly wide variety of substrates (Figure I.5). Counted among the best characterized of these enzymes are 2,3-lysine aminomutase (LAM), involved in lysine metabolism[25,96-98], the activases of Pfl and class III ribonucleotide reductase (aRNR), which generate a stable glycyl radical on their protein substrates[24,99-102], biotin synthase (BioB) and lipoate synthase (LipA), two sulfur insertases[28,35,103-105], and spore photoproduct lyase (SPL), an enzyme capable of repairing

DNA thymidine dimers[106]. Of the characterized AdoMet radical enzymes, LAM and SPL are known to regenerate AdoMet after each turnover[39,107,108], using it as a cofactor rather than cleaving it irreversibly, as observed in the other characterized enzymes[3,39].

### I.C.2. AdoMet radical structure

Since these enzymes were known to generate 5'-dA• through a common mechanism[26,30,33,39], they were also expected to adopt similar folds. Information about the core protein fold required for AdoMet radical activity (generation of 5'-dA• from a $[4Fe-4S]^{1+}$ cluster and AdoMet) is currently available from several sources. These include sequence alignments[84], the six structures of AdoMet radical members (coproporphyrinogen III oxidase (HemN)[109], BioB [105], MoaA, an enzyme involved in molybdenum cofactor biosynthesis[110,111], LAM[98], TYW1, which is involved in biosynthesis of a wyobutosine, a modified tRNA base[112,113] and PflAE[114]), comparisons of those structures[31,36,37], and structure-based sequence alignments[115]. The AdoMet radical fold is a splayed, six-stranded partial TIM barrel, which is extended in each enzyme in order to tailor the individual enzyme to its substrate. Notably, the enzymes structurally characterized to date all use relatively small molecules as substrates, making the structure of the activating enzyme reported in Chapter III highly anticipated.

### I.C.3. The pyruvate formate-lyase system

Decades of effort have led to our current understanding of pyruvate formate-lyase (Pfl) and its mechanism, reactivity and structure[101,116]. Though many questions remain concerning certain aspects of its mechanism and regulation, Pfl is a well characterized system and, along with aRNR, is considered one of the two prototypical Glycyl Radical

Enzymes (GRE, see below)[117,118]. Through the use of an anaerobically stable radical harbored by an active site glycine residue ($G_{734}$), Pfl catalyzes the fermentation of pyruvate to acetyl-CoA and formate via a homolytic carbon-carbon bond cleavage reaction (Figure I.6)[101]. This enzyme, found in facultative and obligate anaerobes, serves as the organism's sole source of acetyl-CoA for the Krebs cycle under fermentative conditions[100].

Because it catalyzes such an important metabolic reaction, several mechanisms of regulation are employed to control Pfl[100,119-123]. In addition to transcriptional regulation, Pfl requires posttranslational activation, which results in introduction of a radical on $G_{734}$[24,99-101]. The activation reaction occurs via stereospecific hydrogen atom abstraction and depends upon PflAE (Figure I.7)[99,124]. Because exposure of the activated Pfl to oxygen results in irreversible cleavage of the polypeptide chain at the radical-harboring glycine[101], a mechanism exists to restore Pfl activity in case of brief oxygen exposure. A short protein with homology to the last 59 residues of Pfl, including $G_{734}$, associates with cleaved Pfl and can interact with PflAE to yield an active enzyme complex[125]. This short protein, called YfiD in *E. coli*, is thought to act as a "spare part" to repair oxygen-cleaved Pfl with minimal energy expenditure[125]. Little is known at present about these mechanisms of posttranslational regulation of Pfl, and research is currently ongoing.

Multiple structures of the inactive form of Pfl have described it as a ten-stranded barrel made up of two five-stranded β sheets in an antiparallel orientation with respect to each other[126-131]. The active site is located at the center of the barrel, made up of residues from the barrel and from two "finger loops" that extend into the barrel from either side (Figure I.8). The radical-harboring glycine residue is always found at the tip of one of these two finger loops, buried far from the surface (8 Å from the surface of Pfl[126]).

Many similarities between Pfl and aRNR exist, including their use of a glycyl radical, their similar structures and their reliance on an activating enzyme[24,99-102]. Since the discovery and characterizations of these two enzymes, several other related enzymes have been identified[132-135] that seem to share these features as well. This family of enzymes has been referred to as the Glycyl Radical Enzymes (GREs)[117,118,134], and at present includes Pfl, aRNR, glycerol dehydratase[133] (Gdh), benzylsuccinate synthase[132] (Bss) and 4-hydroxyphenylacetate decarboxylase[135] (Hpd). Each of these enzymes requires introduction of a glycyl radical by an AdoMet radical activating enzyme, but the details of that activation reaction are not currently understood.

*I.C.4. Remaining questions concerning the AdoMet radical superfamily*

The basic radical initiation reaction held in common by the AdoMet radical enzymes and outlined in Figure I.9 is now generally accepted. The field is now poised to take advantage of the wealth of information the laboratories of Frey, Marquet, Jarrett, Broderick, Booker, Fontecave, Hoffman, Cronan and many others have provided, to fill the remaining gaps in our understanding of this superfamily and then extend that knowledge to other family members, the majority of which are completely uncharacterized. With the structural analyses of several superfamily members complete, we can now focus on obtaining a more detailed understanding of the individual AdoMet radical enzymes and answering the remaining questions alluded to in the above sections and summarized here. How exactly do these enzymes surmount the barrier to AdoMet cleavage presented by the large differences in reduction potential of AdoMet and a typical 4Fe-4S cluster? Can we more clearly describe the inner-sphere electron transfer we envision occurring between the 4Fe-4S cluster and the AdoMet sulfonium to affect

cleavage of AdoMet? Do members of the AdoMet radical superfamily share a specific mechanism of control over the 5'-dA• intermediate, or does this vary along with the different reactions each individual enzyme catalyzes and substrates they bind? In addition, the specific factors that govern AdoMet usage as cofactor or cosubstrate are still unclear, even with a range of AdoMet radical structures in hand. Related to this, how one enzyme is capable of using two molecules of AdoMet during the course of one turnover is controversial, as we will discuss in more detail in Chapter IV. Finally, perhaps the most far-reaching unknown regarding the AdoMet radical family concerns plasticity of the fold; for example, how is the same basic core fold modified to enable so many different types of chemistry, what are the detailed mechanisms of each enzyme after radical generation and how does this impact our understanding of the evolution of the superfamily?

Perhaps with the exception of exceedingly well-characterized LAM, so many details have yet to be understood about the reactions catalyzed by the individual AdoMet radical enzymes. The process of cluster reconstitution in the BioB and LipA systems to allow multiple turnovers, if it does indeed occur, has yet to be shown *in vitro*[35,136]. Definition of the requirements for full activity of many AdoMet radical enzymes such as BioB, LipA, HemN and MoaA is a fundamental problem plaguing the field that has yet to be resolved[34,110,137,138]. Very little is known regarding how glycyl radical formation occurs in the GRE / GRE activase systems; specifically, complete characterization of the proposed conformational change is an important and challenging next step[124,129]. Finally, sequence alignments show that although the Class III RNR activase catalyzes the same reaction as the other GRE activases, it likely differs from these enzymes in terms of its overall fold and residues in the active site[115]. Further characterization of this particular

24

enzyme will assuredly yield additional insight into plasticity and evolution of the AdoMet radical fold.

The work presented in Chapters III and IV of this thesis aims to help answer questions concerning catalysis by the GRE activases, and in particular, PflAE. In Chapter III, we describe the structure of PflAE in both the apo and substrate-bound forms, revealing the first structure of an AdoMet radical activase. We will discuss the architecture of PflAE and its active site, providing a structural basis for understanding glycyl radical formation by this enzyme. In addition, we describe docking studies conducted between PflAE and a fragment of the Pfl structure in order to provide a testable hypothesis for full complex formation between the two enzymes. In the context of the AdoMet radical core fold, the significance of these structures lies in the fact that PflAE is the most basic AdoMet radical enzyme structurally characterized to date. Then, in Chapter IV, we will examine the structures of PflAE, HemN, BioB, MoaA and LAM in detail, comparing them in terms of their overall architecture, AdoMet binding site and substrate binding sites, and discussing the implications of their similarities and differences on catalysis.

## I.D. References

1       Beinert, H., Holm, R. H., and Munck, E., Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* 277 (5326), 653 (1997).

2       Beinert, H., Iron-sulfur proteins: ancient structures, still full of surprises. *J Biol Inorg Chem* 5 (1), 2 (2000).

3       Marquet, A., Tse Sum Bui, B., Smith, A. G., and Warren, M. J., Iron-sulfur proteins as initiators of radical chemistry. *Nat Prod Rep* 24 (5), 1027 (2007).

4       Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K., Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* 74, 247 (2005).

[5] Brzoska, K., Meczynska, S., and Kruszewski, M., Iron-sulfur cluster proteins: electron transfer and beyond. *Acta Biochim Pol* **53** (4), 685 (2006).

[6] Rees, D. C., Great metalloclusters in enzymology. *Annu Rev Biochem* **71**, 221 (2002).

[7] Drennan, C. L. and Peters, J. W., Surprising cofactors in metalloenzymes. *Curr Opin Struct Biol* **13** (2), 220 (2003).

[8] Drennan, C. L., Doukov, T. I., and Ragsdale, S. W., The metalloclusters of carbon monoxide dehydrogenase/acetyl-CoA synthase: a story in pictures. *J Biol Inorg Chem* **9** (5), 511 (2004).

[9] Kim, J. and Rees, D. C., Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science* **257** (5077), 1677 (1992).

[10] Kim, J., Woo, D., and Rees, D. C., X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. *Biochemistry* **32** (28), 7104 (1993).

[11] Einsle, O. et al., Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor. *Science* **297** (5587), 1696 (2002).

[12] Peters, J. W., Lanzilotta, W. N., Lemon, B. J., and Seefeldt, L. C., X-ray crystal structure of the Fe-only hydrogenase (CpI) from Clostridium pasteurianum to 1.8 angstrom resolution. *Science* **282** (5395), 1853 (1998).

[13] Nicolet, Y. et al., Desulfovibrio desulfuricans iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* **7** (1), 13 (1999).

[14] Drennan, C. L. et al., Life on carbon monoxide: X-ray structure of Rhodospirillum rubrum Ni-Fe-S carbon monoxide dehydrogenase. *Proc Natl Acad Sci U S A* **98** (21), 11973 (2001).

[15] Dobbek, H. et al., Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* **293** (5533), 1281 (2001).

[16] Doukov, T. I. et al., A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* **298** (5593), 567 (2002).

[17] Lill, R. et al., The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. *Biol Chem* **380** (10), 1157 (1999).

[18] Balk, J. and Lill, R., The cell's cookbook for iron--sulfur clusters: recipes for fool's gold? *Chembiochem* **5** (8), 1044 (2004).

[19] Barras, F., Loiseau, L., and Py, B., How Escherichia coli and Saccharomyces cerevisiae build Fe/S proteins. *Adv Microb Physiol* **50**, 41 (2005).

[20] Lill, R. and Muhlenhoff, U., Iron-sulfur-protein biogenesis in eukaryotes. *Trends Biochem Sci* **30** (3), 133 (2005).

[21] Lill, R. and Muhlenhoff, U., Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. *Annu Rev Cell Dev Biol* **22**, 457 (2006).

[22] Pilon, M. et al., Biogenesis of iron-sulfur cluster proteins in plastids. *Genet Eng (N Y)* **27**, 101 (2006).

[23] Broderick, J. B., Assembling iron-sulfur clusters in the cytosol. *Nat Chem Biol* **3** (5), 243 (2007).

[24] Kozarich, J. W., S-adenosylmethionine-dependent enzyme activation. *Biofactors* **1** (2), 123 (1988).

[25] Frey, P. A., Lysine 2,3-aminomutase: is adenosylmethionine a poor man's adenosylcobalamin? *FASEB J* **7** (8), 662 (1993).

[26]  Cheek, J. and Broderick, J. B., Adenosylmethionine-dependent iron-sulfur enzymes: versatile clusters in a radical new role. *J Biol Inorg Chem* **6** (3), 209 (2001).

[27]  Fontecave, M., Mulliez, E., and Ollagnier-de-Choudens, S., Adenosylmethionine as a source of 5'-deoxyadenosyl radicals. *Curr Opin Chem Biol* **5** (5), 506 (2001).

[28]  Frey, P. A. and Booker, S. J., Radical mechanisms of S-adenosylmethionine-dependent enzymes. *Adv Protein Chem* **58**, 1 (2001).

[29]  Frey, P. A., The role of radicals in enzymatic processes. *Chem Rec* **1** (4), 277 (2001).

[30]  Jarrett, J. T., The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. *Curr Opin Chem Biol* **7** (2), 174 (2003).

[31]  Layer, G., Heinz, D. W., Jahn, D., and Schubert, W. D., Structure and function of radical SAM enzymes. *Curr Opin Chem Biol* **8** (5), 468 (2004).

[32]  Marsh, E. N., Patwardhan, A., and Huhta, M. S., S-adenosylmethionine radical enzymes. *Bioorg Chem* **32** (5), 326 (2004).

[33]  Buis, J. M. and Broderick, J. B., Pyruvate formate-lyase activating enzyme: elucidation of a novel mechanism for glycyl radical formation. *Arch Biochem Biophys* **433** (1), 288 (2005).

[34]  Jarrett, J. T., The novel structure and chemistry of iron-sulfur clusters in the adenosylmethionine-dependent radical enzyme biotin synthase. *Arch Biochem Biophys* **433** (1), 312 (2005).

[35]  Lotierzo, M. et al., Biotin synthase mechanism: an overview. *Biochem Soc Trans* **33** (Pt 4), 820 (2005).

[36]  Layer, G. et al., Structural and functional comparison of HemN to other radical SAM enzymes. *Biol Chem* **386** (10), 971 (2005).

[37]  Grillo, M. A. and Colombatto, S., S-adenosylmethionine and radical-based catalysis. *Amino Acids* **32** (2), 197 (2007).

[38]  Roje, S., S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* **67** (15), 1686 (2006).

[39]  Wang, S. C. and Frey, P. A., S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem Sci* **32** (3), 101 (2007).

[40]  Booker, S. J., Cicchillo, R. M., and Grove, T. L., Self-sacrifice in radical S-adenosylmethionine proteins. *Curr Opin Chem Biol* (2007).

[41]  Zheng, L. et al., Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc Natl Acad Sci U S A* **90** (7), 2754 (1993).

[42]  Kessler, D., Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes. *FEMS Microbiol Rev* **30** (6), 825 (2006).

[43]  Mendel, R. R., Smith, A. G., Marquet, A., and Warren, M. J., Metal and cofactor insertion. *Nat Prod Rep* **24** (5), 963 (2007).

[44]  Holliday, G. L. et al., Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep* **24** (5), 972 (2007).

[45]  Djaman, O., Outten, F. W., and Imlay, J. A., Repair of oxidized iron-sulfur clusters in Escherichia coli. *J Biol Chem* **279** (43), 44590 (2004).

[46]  Fontecave, M., Choudens, S. O., Py, B., and Barras, F., Mechanisms of iron-sulfur cluster assembly: the SUF machinery. *J Biol Inorg Chem* **10** (7), 713 (2005).

[47] Xu, X. M. and Moller, S. G., AtNAP7 is a plastidic SufC-like ATP-binding cassette/ATPase essential for Arabidopsis embryogenesis. *Proc Natl Acad Sci U S A* **101** (24), 9143 (2004).

[48] Tirupati, B., Vey, J. L., Drennan, C. L., and Bollinger, J. M., Jr., Kinetic and structural characterization of Slr0077/SufS, the essential cysteine desulfurase from Synechocystis sp. PCC 6803. *Biochemistry* **43** (38), 12210 (2004).

[49] Xu, X. M., Adams, S., Chua, N. H., and Moller, S. G., AtNAP1 represents an atypical SufB protein in Arabidopsis plastids. *J Biol Chem* **280** (8), 6648 (2005).

[50] Vickery, L. E. and Cupp-Vickery, J. R., Molecular chaperones HscA/Ssq1 and HscB/Jac1 and their roles in iron-sulfur protein maturation. *Crit Rev Biochem Mol Biol* **42** (2), 95 (2007).

[51] Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N., Kinetic and mutational studies of three NifS homologs from Escherichia coli: mechanistic difference between L-cysteine desulfurase and L-selenocysteine lyase reactions. *J Biochem (Tokyo)* **127** (4), 559 (2000).

[52] Behshad, E., Parkin, S. E., and Bollinger, J. M., Jr., Mechanism of cysteine desulfurase Slr0387 from Synechocystis sp. PCC 6803: kinetic analysis of cleavage of the persulfide intermediate by chemical reductants. *Biochemistry* **43** (38), 12220 (2004).

[53] Loiseau, L. et al., Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J Biol Chem* **278** (40), 38352 (2003).

[54] Ollagnier-de-Choudens, S. et al., Mechanistic studies of the SufS-SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE. *FEBS Lett* **555** (2), 263 (2003).

[55] Outten, F. W., Wood, M. J., Munoz, F. M., and Storz, G., The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in Escherichia coli. *J Biol Chem* **278** (46), 45713 (2003).

[56] Ding, H. and Clark, R. J., Characterization of iron binding in IscA, an ancient iron-sulphur cluster assembly protein. *Biochem J* **379** (Pt 2), 433 (2004).

[57] Ding, H., Clark, R. J., and Ding, B., IscA mediates iron delivery for assembly of iron-sulfur clusters in IscU under the limited accessible free iron conditions. *J Biol Chem* **279** (36), 37499 (2004).

[58] Lu, J., Yang, J., Tan, G., and Ding, H., Complementary roles of SufA and IscA in the biogenesis of iron-sulfur clusters in Escherichia coli. *Biochem J* (2007).

[59] Hoff, K. G., Silberg, J. J., and Vickery, L. E., Interaction of the iron-sulfur cluster assembly protein IscU with the Hsc66/Hsc20 molecular chaperone system of Escherichia coli. *Proc Natl Acad Sci U S A* **97** (14), 7790 (2000).

[60] Hoff, K. G., Cupp-Vickery, J. R., and Vickery, L. E., Contributions of the LPPVK motif of the iron-sulfur template protein IscU to interactions with the Hsc66-Hsc20 chaperone system. *J Biol Chem* **278** (39), 37582 (2003).

[61] Tapley, T. L. and Vickery, L. E., Preferential substrate binding orientation by the molecular chaperone HscA. *J Biol Chem* **279** (27), 28435 (2004).

[62] Nachin, L., Loiseau, L., Expert, D., and Barras, F., SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *EMBO J* **22** (3), 427 (2003).

63   Nachin, L. et al., SoxR-dependent response to oxidative stress and virulence of Erwinia chrysanthemi: the key role of SufC, an orphan ABC ATPase. *Mol Microbiol* **39** (4), 960 (2001).

64   Schwartz, C. J. et al., IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins. *Proc Natl Acad Sci U S A* **98** (26), 14895 (2001).

65   Outten, F. W., Djaman, O., and Storz, G., A suf operon requirement for Fe-S cluster assembly during iron starvation in Escherichia coli. *Mol Microbiol* **52** (3), 861 (2004).

66   Lee, J. H., Yeo, W. S., and Roe, J. H., Induction of the sufA operon encoding Fe-S assembly proteins by superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-responsive factor. *Mol Microbiol* **51** (6), 1745 (2004).

67   Wang, T. et al., The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria. *J Bacteriol* **186** (4), 956 (2004).

68   Seki, A. et al., Light-responsive transcriptional regulation of the suf promoters involved in cyanobacterium Synechocystis sp. PCC 6803 Fe-S cluster biogenesis. *FEBS Lett* **580** (21), 5044 (2006).

69   Shen, G. et al., SufR coordinates two [4Fe-4S]2+,1+ clusters and functions as a transcriptional repressor of the sufBCDS operon and an autoregulator of sufR in cyanobacteria. *J Biol Chem* (2007).

70   Yoon, T. and Cowan, J. A., Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins. *J Am Chem Soc* **125** (20), 6078 (2003).

71   Ollagnier-de-Choudens, S., Sanakis, Y., and Fontecave, M., SufA/IscA: reactivity studies of a class of scaffold proteins involved in [Fe-S] cluster assembly. *J Biol Inorg Chem* **9** (7), 828 (2004).

72   Krebs, C. et al., IscA, an alternate scaffold for Fe-S cluster biosynthesis. *Biochemistry* **40** (46), 14069 (2001).

73   Ollagnier-de-Choudens, S., Mattioli, T., Takahashi, Y., and Fontecave, M., Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J Biol Chem* **276** (25), 22604 (2001).

74   Ding, H., Yang, J., Coleman, L. C., and Yeung, S., Distinct iron binding property of two putative iron donors for the iron-sulfur cluster assembly: IscA and the bacterial frataxin ortholog CyaY under physiological and oxidative stress conditions. *J Biol Chem* **282** (11), 7997 (2007).

75   Mansy, S. S., Wu, G., Surerus, K. K., and Cowan, J. A., Iron-sulfur cluster biosynthesis. Thermatoga maritima IscU is a structured iron-sulfur cluster assembly protein. *J Biol Chem* **277** (24), 21397 (2002).

76   Bertini, I. et al., Thermotoga maritima IscU. Structural characterization and dynamics of a new class of metallochaperone. *J Mol Biol* **331** (4), 907 (2003).

77   Mansy, S. S., Wu, S. P., and Cowan, J. A., Iron-sulfur cluster biosynthesis: biochemical characterization of the conformational dynamics of Thermotoga maritima IscU and the relevance for cellular cluster assembly. *J Biol Chem* **279** (11), 10469 (2004).

[78] Bilder, P. W., Ding, H., and Newcomer, M. E., Crystal structure of the ancient, Fe-S scaffold IscA reveals a novel protein fold. *Biochemistry* **43** (1), 133 (2004).

[79] Adinolfi, S. et al., Bacterial IscU is a well folded and functional single domain protein. *Eur J Biochem* **271** (11), 2093 (2004).

[80] Liu, J. et al., Structural characterization of an iron-sulfur cluster assembly protein IscU in a zinc-bound form. *Proteins* **59** (4), 875 (2005).

[81] Shimomura, Y. et al., Characterization and Crystallization of an IscU-Type Scaffold Protein with Bound [2Fe-2S] Cluster from the Hyperthermophile, Aquifex aeolicus. *J Biochem (Tokyo)* **mvm163** (1) (2007).

[82] Cupp-Vickery, J. R., Peterson, J. C., Ta, D. T., and Vickery, L. E., Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. *J Mol Biol* **342** (4), 1265 (2004).

[83] Tapley, T. L., Cupp-Vickery, J. R., and Vickery, L. E., Structural determinants of HscA peptide-binding specificity. *Biochemistry* **45** (26), 8058 (2006).

[84] Sofia, H. J. et al., Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **29** (5), 1097 (2001).

[85] Walsby, C. J. et al., Electron-nuclear double resonance spectroscopic evidence that S-adenosylmethionine binds in contact with the catalytically active [4Fe-4S](+) cluster of pyruvate formate-lyase activating enzyme. *J Am Chem Soc* **124** (12), 3143 (2002).

[86] Walsby, C. J. et al., An anchoring role for FeS clusters: chelation of the amino acid moiety of S-adenosylmethionine to the unique iron site of the [4Fe-4S] cluster of pyruvate formate-lyase activating enzyme. *J Am Chem Soc* **124** (38), 11270 (2002).

[87] Krebs, C. et al., Coordination of adenosylmethionine to a unique iron site of the [4Fe-4S] of pyruvate formate-lyase activating enzyme: a Mossbauer spectroscopic study. *J Am Chem Soc* **124** (6), 912 (2002).

[88] Cosper, M. M. et al., The [4Fe-4S](2+) cluster in reconstituted biotin synthase binds S-adenosyl-L-methionine. *J Am Chem Soc* **124** (47), 14006 (2002).

[89] Chen, D., Walsby, C., Hoffman, B. M., and Frey, P. A., Coordination and mechanism of reversible cleavage of S-adenosylmethionine by the [4Fe-4S] center in lysine 2,3-aminomutase. *J Am Chem Soc* **125** (39), 11788 (2003).

[90] Walsby, C. J. et al., Spectroscopic approaches to elucidating novel iron-sulfur chemistry in the "radical-Sam" protein superfamily. *Inorg Chem* **44** (4), 727 (2005).

[91] Blaschkowski, H. P., Neuer, G., Ludwig-Festl, M., and Knappe, J., Routes of flavodoxin and ferredoxin reduction in Escherichia coli. CoA-acylating pyruvate: flavodoxin and NADPH: flavodoxin oxidoreductases participating in the activation of pyruvate formate-lyase. *Eur J Biochem* **123** (3), 563 (1982).

[92] Bianchi, V. et al., Flavodoxin is required for the activation of the anaerobic ribonucleotide reductase. *Biochem Biophys Res Commun* **197** (2), 792 (1993).

[93] Birch, O. M., Fuhrmann, M., and Shaw, N. M., Biotin synthase from Escherichia coli, an investigation of the low molecular weight and protein components required for activity in vitro. *J Biol Chem* **270** (32), 19158 (1995).

[94] Brazeau, B. J. et al., Enzymatic activation of lysine 2,3-aminomutase from Porphyromonas gingivalis. *Appl Environ Microbiol* **72** (9), 6402 (2006).

[95] Cosper, M. M. et al., Structural studies of the interaction of S-adenosylmethionine with the [4Fe-4S] clusters in biotin synthase and pyruvate formate-lyase activating enzyme. *Protein Sci* **12** (7), 1573 (2003).

[96] Chirpich, T. P., Zappia, V., Costilow, R. N., and Barker, H. A., Lysine 2,3-aminomutase. Purification and properties of a pyridoxal phosphate and S-adenosylmethionine-activated enzyme. *J Biol Chem* **245** (7), 1778 (1970).

[97] Frey, P. A., Ballinger, M. D., and Reed, G. H., S-adenosylmethionine: a 'poor man's coenzyme B12' in the reaction of lysine 2,3-aminomutase. *Biochem Soc Trans* **26** (3), 304 (1998).

[98] Lepore, B. W., Ruzicka, F. J., Frey, P. A., and Ringe, D., The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale. *Proc Natl Acad Sci U S A* **102** (39), 13819 (2005).

[99] Conradt, H. et al., Pyruvate formate-lyase (inactive form) and pyruvate formate-lyase activating enzyme of Escherichia coli: isolation and structural properties. *Arch Biochem Biophys* **228** (1), 133 (1984).

[100] Knappe, J. and Sawers, G., A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of Escherichia coli. *FEMS Microbiol Rev* **6** (4), 383 (1990).

[101] Knappe, J. and Wagner, A. F., Glycyl free radical in pyruvate formate-lyase: synthesis, structure characteristics, and involvement in catalysis. *Methods Enzymol* **258**, 343 (1995).

[102] Tamarit, J. et al., The anaerobic ribonucleotide reductase from Escherichia coli. The small protein is an activating enzyme containing a [4fe-4s](2+) center. *J Biol Chem* **274** (44), 31291 (1999).

[103] Hayden, M. A. et al., Biosynthesis of lipoic acid: characterization of the lipoic acid auxotrophs Escherichia coli W1485-lip2 and JRG33-lip9. *Biochemistry* **32** (14), 3778 (1993).

[104] Miller, J. R. et al., Escherichia coli LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein. *Biochemistry* **39** (49), 15166 (2000).

[105] Berkovitch, F. et al., Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* **303** (5654), 76 (2004).

[106] Rebeil, R. et al., Spore photoproduct lyase from Bacillus subtilis spores is a novel iron-sulfur DNA repair enzyme which shares features with proteins such as class III anaerobic ribonucleotide reductases and pyruvate-formate lyases. *J Bacteriol* **180** (18), 4879 (1998).

[107] Moss, M. and Frey, P. A., The role of S-adenosylmethionine in the lysine 2,3-aminomutase reaction. *J Biol Chem* **262** (31), 14859 (1987).

[108] Cheek, J. and Broderick, J. B., Direct H atom abstraction from spore photoproduct C-6 initiates DNA repair in the reaction catalyzed by spore photoproduct lyase: evidence for a reversibly generated adenosyl radical intermediate. *J Am Chem Soc* **124** (12), 2860 (2002).

[109] Layer, G. et al., Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J* **22** (23), 6214 (2003).

110 Hanzelmann, P. and Schindelin, H., Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc Natl Acad Sci U S A* **101** (35), 12870 (2004).

111 Hanzelmann, P. and Schindelin, H., Binding of 5'-GTP to the C-terminal FeS cluster of the radical S-adenosylmethionine enzyme MoaA provides insights into its mechanism. *Proc Natl Acad Sci U S A* **103** (18), 6829 (2006).

112 Suzuki, Y. et al., Crystal Structure of the Radical SAM Enzyme Catalyzing Tricyclic Modified Base Formation in tRNA. *J Mol Biol* **372** (5), 1204 (2007).

113 Goto-Ito, S. et al., Structure of an archaeal TYW1, the enzyme catalyzing the second step of wye-base biosynthesis. *Acta Crystallogr D Biol Crystallogr* **63** (Pt 10), 1059 (2007).

114 Vey, J. Yang, J., Li, M., Broderick, W., Broderick, J., Drennan, C.

115 Nicolet, Y. and Drennan, C. L., AdoMet radical proteins--from structure to evolution--alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* **32** (13), 4015 (2004).

116 Knappe, J., Elbert, S., Frey, M., and Wagner, A. F., Pyruvate formate-lyase mechanism involving the protein-based glycyl radical. *Biochem Soc Trans* **21** (3), 731 (1993).

117 Eklund, H. and Fontecave, M., Glycyl radical enzymes: a conservative structural basis for radicals. *Structure* **7** (11), R257 (1999).

118 Buckel, W. and Golding, B. T., Radical enzymes in anaerobes. *Annu Rev Microbiol* **60**, 27 (2006).

119 Sawers, G. and Bock, A., Anaerobic regulation of pyruvate formate-lyase from Escherichia coli K-12. *J Bacteriol* **170** (11), 5330 (1988).

120 Sawers, G. and Bock, A., Novel transcriptional control of the pyruvate formate-lyase gene: upstream regulatory sequences and multiple promoters regulate anaerobic expression. *J Bacteriol* **171** (5), 2485 (1989).

121 Sawers, G. and Suppmann, B., Anaerobic induction of pyruvate formate-lyase gene expression is mediated by the ArcA and FNR proteins. *J Bacteriol* **174** (11), 3474 (1992).

122 Sawers, G., Specific transcriptional requirements for positive regulation of the anaerobically inducible pfl operon by ArcA and FNR. *Mol Microbiol* **10** (4), 737 (1993).

123 Nnyepi, M. R., Peng, Y., and Broderick, J. B., Inactivation of E. coli pyruvate formate-lyase: role of AdhE and small molecules. *Arch Biochem Biophys* **459** (1), 1 (2007).

124 Frey, M., Rothe, M., Wagner, A. F., and Knappe, J., Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. *J Biol Chem* **269** (17), 12432 (1994).

125 Wagner, A. F. et al., YfiD of Escherichia coli and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem Biophys Res Commun* **285** (2), 456 (2001).

[126] Becker, A. et al., Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Biol* **6** (10), 969 (1999).

[127] Logan, D. T., Andersson, J., Sjoberg, B. M., and Nordlund, P., A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283** (5407), 1499 (1999).

[128] Leppanen, V. M. et al., Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure* **7** (7), 733 (1999).

[129] Lehtio, L., Leppanen, V. M., Kozarich, J. W., and Goldman, A., Structure of Escherichia coli pyruvate formate-lyase with pyruvate. *Acta Crystallogr D Biol Crystallogr* **58** (Pt 12), 2209 (2002).

[130] O'Brien, J. R. et al., Insight into the mechanism of the B12-independent glycerol dehydratase from Clostridium butyricum: preliminary biochemical and structural characterization. *Biochemistry* **43** (16), 4635 (2004).

[131] Lehtio, L. et al., Crystal structure of a glycyl radical enzyme from Archaeoglobus fulgidus. *J Mol Biol* **357** (1), 221 (2006).

[132] Leuthner, B. et al., Biochemical and genetic characterization of benzylsuccinate synthase from Thauera aromatica: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28** (3), 615 (1998).

[133] Raynaud, C. et al., Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum. *Proc Natl Acad Sci U S A* **100** (9), 5010 (2003).

[134] Selmer, T., Pierik, A. J., and Heider, J., New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* **386** (10), 981 (2005).

[135] Yu, L. et al., 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* **45** (31), 9584 (2006).

[136] Tse Sum Bui, B. et al., Further investigation on the turnover of Escherichia coli biotin synthase with dethiobiotin and 9-mercaptodethiobiotin as substrates. *Biochemistry* **43** (51), 16432 (2004).

[137] Cicchillo, R. M. and Booker, S. J., Mechanistic investigations of lipoic acid biosynthesis in Escherichia coli: both sulfur atoms in lipoic acid are contributed by the same lipoyl synthase polypeptide. *J Am Chem Soc* **127** (9), 2860 (2005).

[138] Layer, G., Verfurth, K., Mahlitz, E., and Jahn, D., Oxygen-independent coproporphyrinogen-III oxidase HemN from Escherichia coli. *J Biol Chem* **277** (37), 34136 (2002).

[139] Kolling, D. J. et al., Atomic resolution structures of rieske iron-sulfur protein: role of hydrogen bonds in tuning the redox potential of iron-sulfur clusters. *Structure* **15** (1), 29 (2007).

[140] Liu, L. et al., Ultrahigh-resolution structure of high-potential iron-sulfur protein from Thermochromatium tepidum. *Acta Crystallogr D Biol Crystallogr* **58** (Pt 7), 1085 (2002).

[141] Peters, J. W. et al., Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36** (6), 1181 (1997).

## I.E. Figures

**Figure I.1: Various known FeS cluster forms**. The FeS clusters and their ligands are shown in ball and stick, colored as follows: protein carbons, grey; oxygen, red; nitrogen, blue; sulfur, gold; iron, ruby; molybdenum, magenta; nickel, teal; copper, green. The clusters shown are: (a) 2Fe-2S cluster from Rieske FeS protein (2NWF)[139] from *Rhodobacter sphaeroides* (b) 4Fe-4S cluster of high-potential FeS protein (1IUA)[140] from *Thermochromatium tepidum* (c) H-cluster of *Desulfovibrio desulfuricans* hydrogenase (1HFE)[13] (d) nitrogenase P-cluster from *Azotobacter vinelandii* in the reduced state (3MIN)[141] (e) FeMo-cofactor from *Azotobacter vinelandii* (PDB 1MIN)[11] (f) The C-cluster of bifunctional CODH/ACS from *Clostridium thermoaceticum* (1MJG)[16] and (g) the A-cluster of CODH/ACS from *Clostridium thermoaceticum* (1MJG)[16]. The copper atom in this cluster binds in a heterogeneous site capable of binding copper, nickel, and zinc.

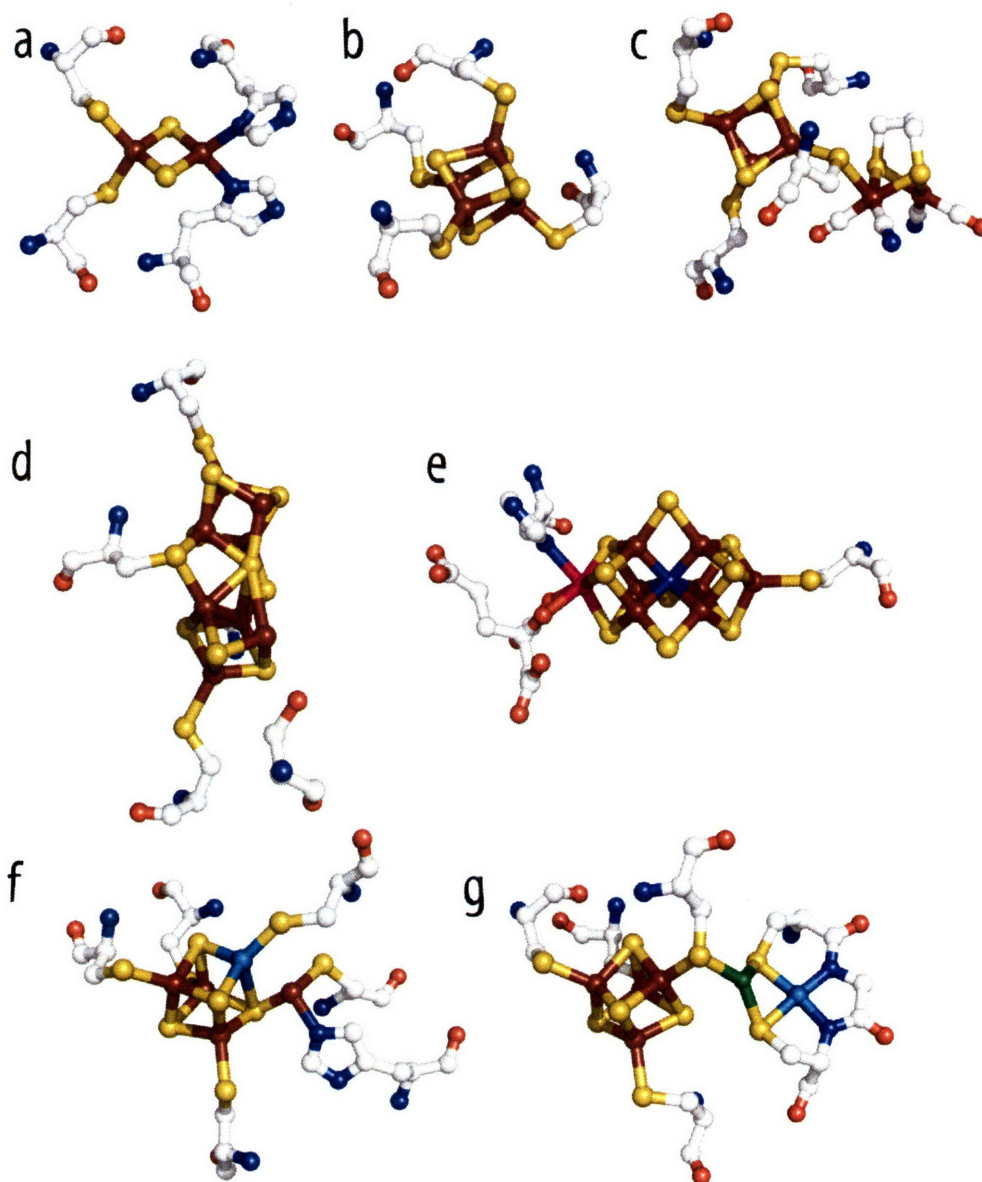**Figure I.2: Schematic of the (a) *isc* and (b) *suf* operons from *E. coli* involved in FeS cluster bioassembly**. The *nif* operon is similar to the *isc* gene cluster. The five proteins marked as "unknown" likely provide energy for the system by a mechanism that is currently uncharacterized. The proteins involved in regulation of the *suf* operon (Fur, OxyR in *E. coli*) are not shown. This figure was adapted from reference 46.
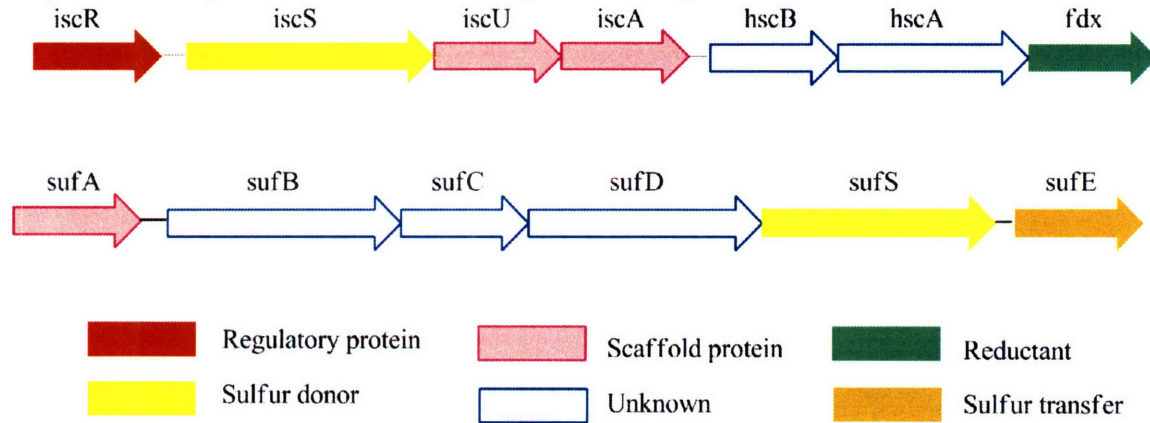
**Figure I.3: The general cluster bioassembly pathway**, as described in reference 46. Colors are the same as in Figure I.2. Briefly, a cysteine desulfurase (with the help of SufE in some systems) and an unidentified iron donor build a cluster into a scaffold protein by transferring sulfur and iron, respectively. The scaffold protein then transfers the cluster to a target protein (white circles). The remaining proteins in each operon – ferredoxin, HscA / HscB, SufBCD – are involved in this process in some way, which at this point is unclear.

apo scaffold protein

[Fe-S]

[  ]

SufE

SufS

holo target protein

?

Iron donor?

?

chaperone/
other proteins

fdx

[  ]

[Fe-S]

SufE

apo target protein

holo scaffold protein

SufS
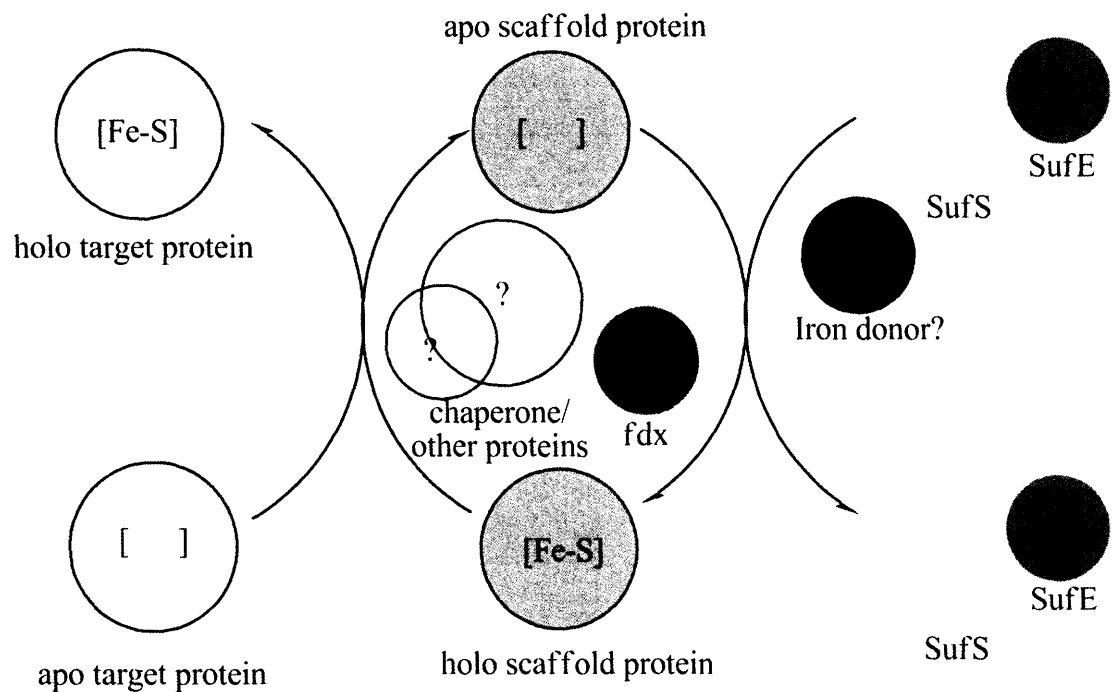
**Figure I.4:  Ligation of the 4Fe-4S cluster by AdoMet in the AdoMet radical enzymes.**  This figure was adapted from reference 90.
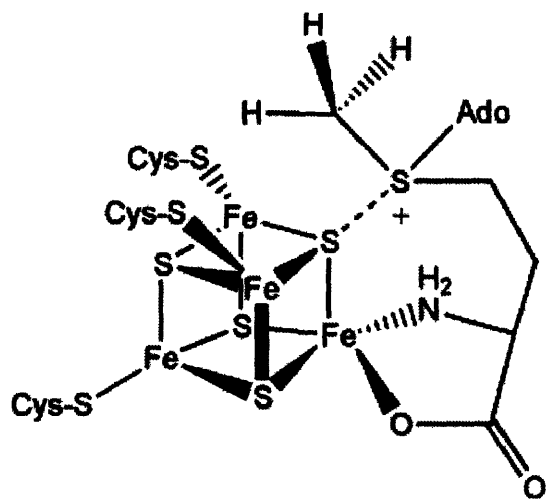
**Figure I.5: Representative AdoMet radical enzyme reactions.** In (b), $R_1$ and $R_2$ correspond to the remainder of the coproporphyrinogen III tetrapyrrole macrocycle. This figure was adapted from reference 39.
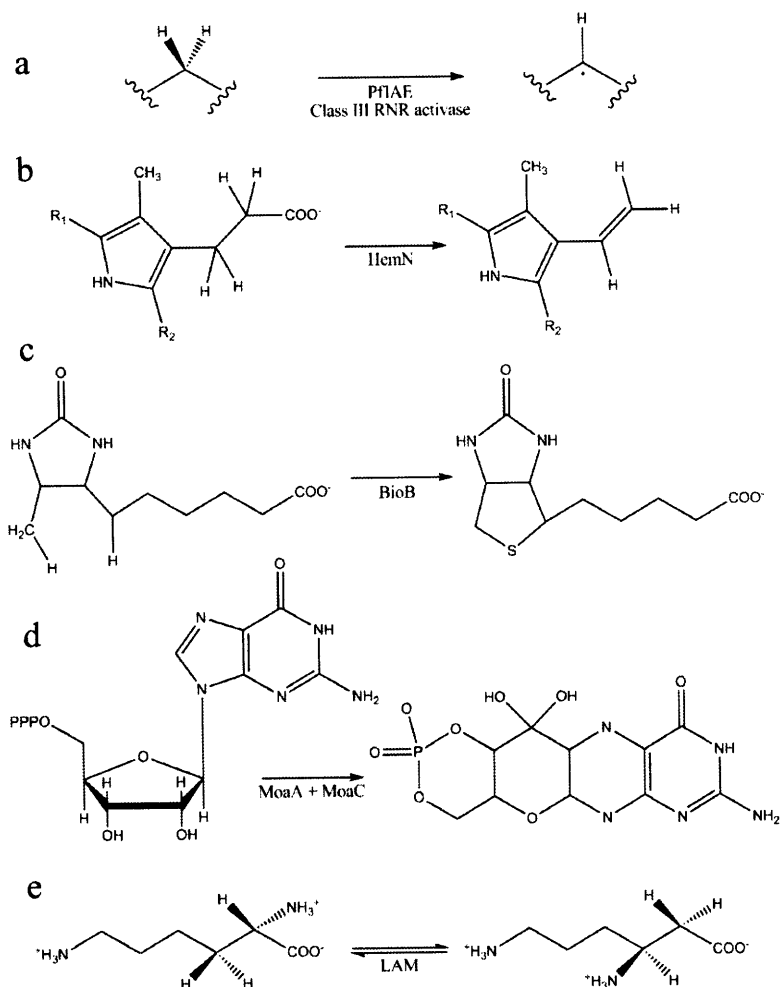
**Figure I.6: Reaction catalyzed by Pfl.** (a) Overall Pfl reaction and (b) Pfl reaction
mechanism, as proposed in reference 126.
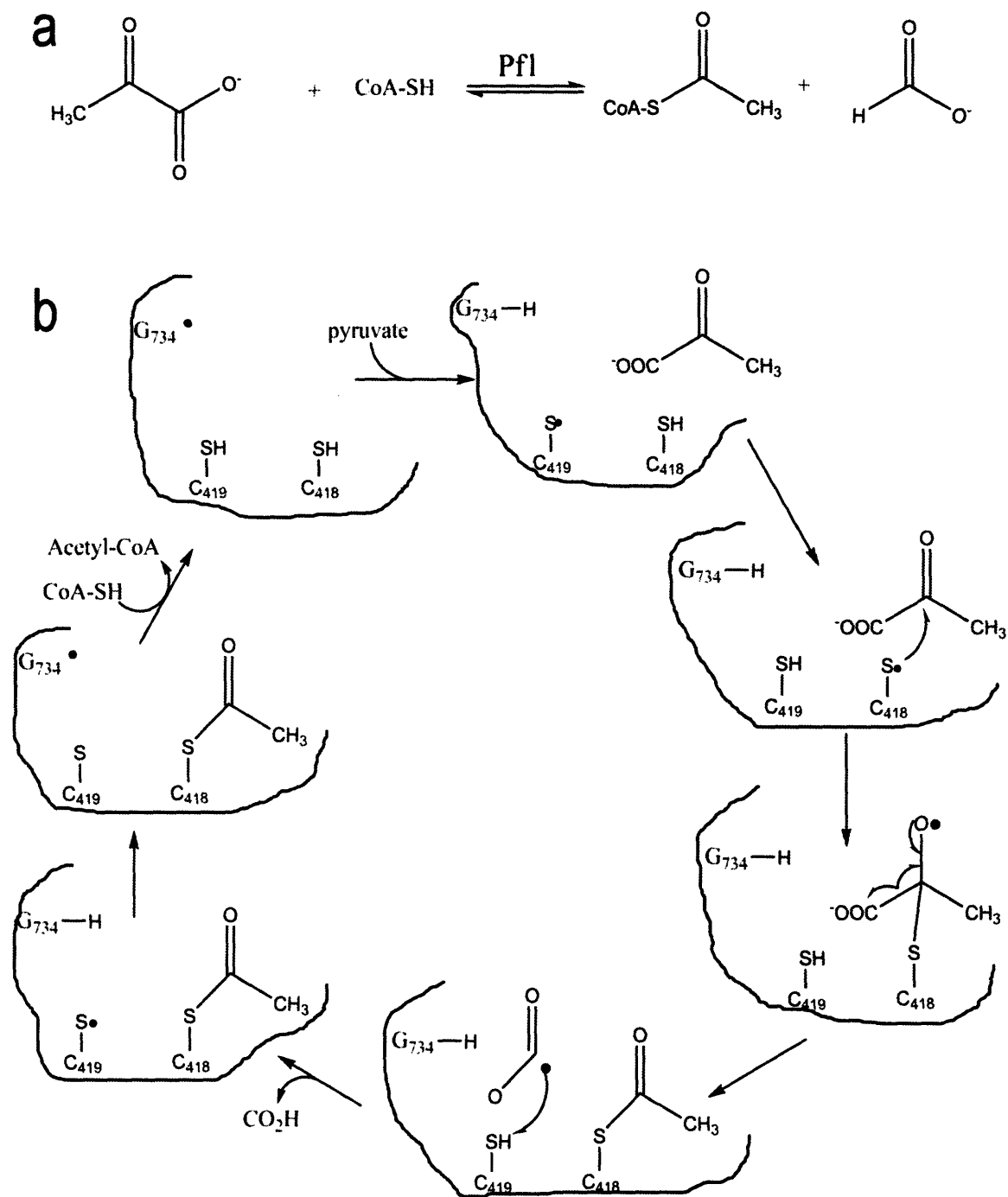
**Figure I.7: Pfl activation reaction catalyzed by PflAE.** This figure was adapted from reference 90.
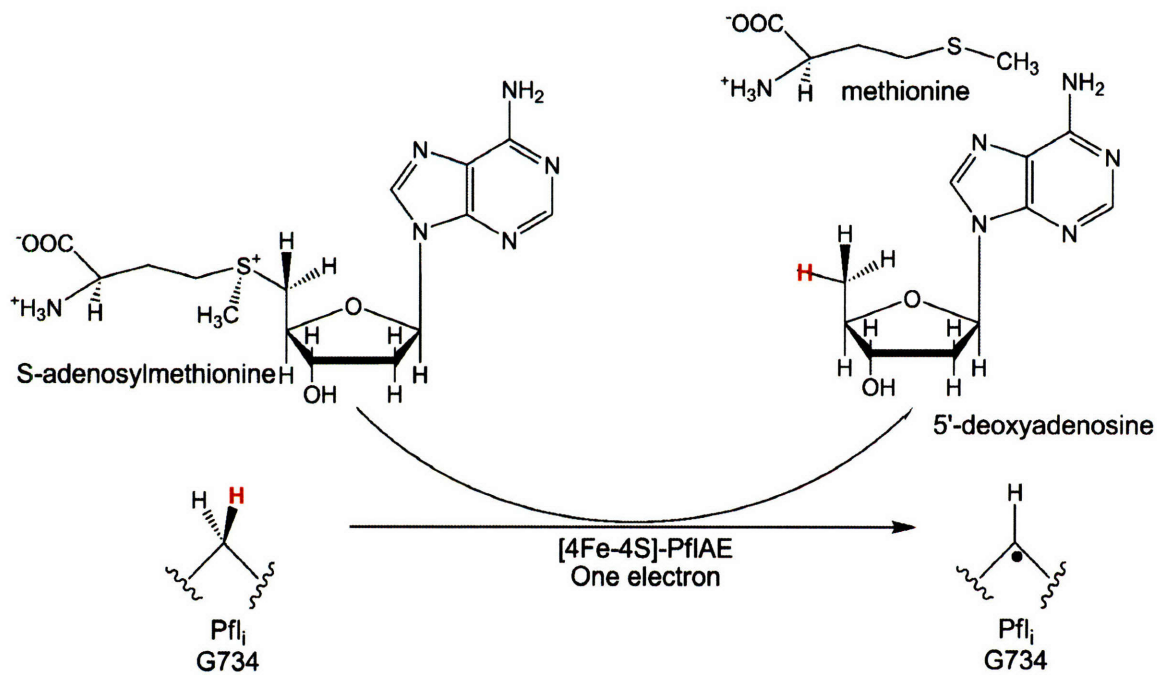
**Figure I.8: Structure of the Pfl monomer.** β strands are colored yellow, α helices magenta, and the two active site finger loops are colored blue (cysteine loop) and red (glycine loop, with $G_{734}$ in spheres).
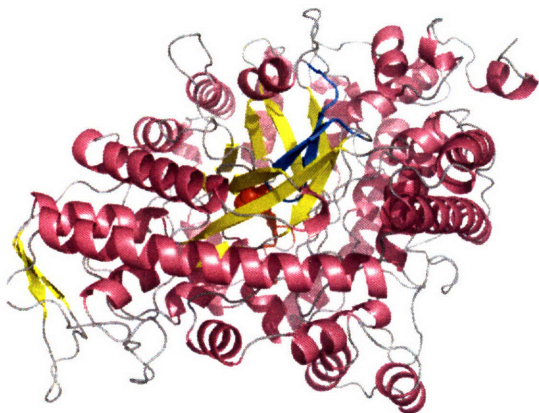


**Figure I.9: Radical generation by the AdoMet radical enzymes.**



S-adenosylmethionine

5'-deoxyadenosyl radical

$[4Fe-4S]^{1+}$-AdoMet radical enzyme

methionine
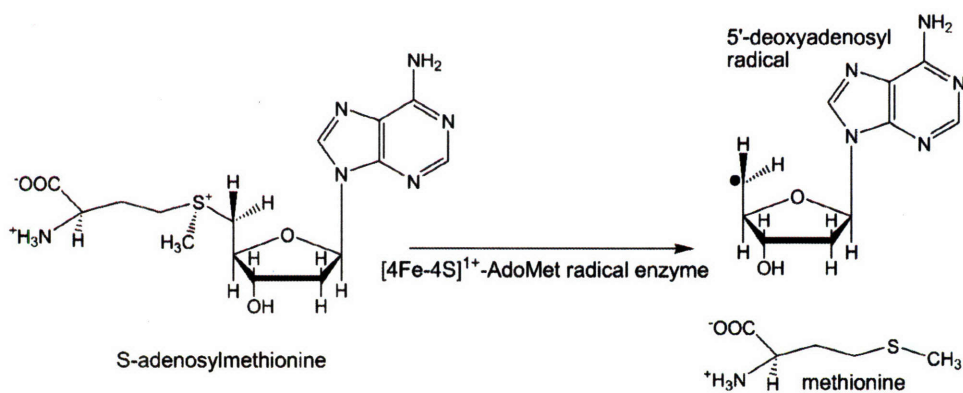
# Chapter II:  Crystal Structure of slr0077/SufS

## II.A. Summary

Cysteine desulfurases play an important role in sulfur metabolism, acting as a source of sulfur for insertion into iron-sulfur clusters, vitamins, cofactors, and other biomolecules.  These enzymes catalyze the $\beta$-elimination of sulfur from a substrate cysteine molecule using pyridoxal 5'-phosphate and an enzyme cysteine sidechain.  Two different types of cysteine desulfurases catalyze this reaction: the Group I IscS-like enzymes, and the Group II SufS-like enzymes.  In this study, the crystal structure of SufS from *Synechocystis* sp. PCC 6803 was solved by molecular replacement with the *E. coli* SufS enzyme as the search model, revealing a structure homologous to the other cysteine desulfurases and identifying a possible binding site for activating factors.

## II.B. Introduction

Sulfur metabolism has been increasingly recognized as an important and complicated process. Biosynthetic pathways that involve incorporation of sulfur into a molecule or cofactor include those of iron-sulfur clusters[2-11], the thiamin[12,13], molybdopterin[14], biotin and lipoate cofactors, and thionucleosides[15-17]. Cysteine (or cystine) is typically used as the source of sulfur for these pathways, which use the persulfidic (or sulfane) form of sulfur during transfer[18]. Cysteine itself is formed directly from the reaction of sulfide with O-acetylserine, demonstrating that a certain amount of free sulfide must be present in cells[19]; however, it is thought that persulfidic sulfur, though as reactive as sulfide and possibly toxic in its own right[20], is more easily handled and therefore better suited for these biosynthetic pathways[18]. Persulfidic sulfur is always found either bound to a protein or as part of a larger molecule, such as thiosulfate, making possible more specific interactions and control.

There are several enzymes known to be capable of forming persulfidic sulfur from cysteine (or related molecules, such as cystine or thiosulfate)[18]. The cysteine desulfurases, NifS, IscS and SufS, are currently the best characterized of these enzymes. Cystine lyase (C-DES)[21,22], cysteine desulfidase[23], rhodanese (and rhodanese-like proteins)[24] and mercaptopyruvate sulfurtransferase[25] can also generate persulfidic sulfur, and have been implicated in some of the biosynthetic pathways mentioned above[26] – in particular, iron-sulfur cluster bioassembly. A key theme observed in the literature on sulfur metabolism (and iron-sulfur cluster biosynthesis, as well) is redundancy, underscoring the importance of these systems and complicating more detailed analyses of individual enzymes and pathways.

As mentioned above, cysteine desulfurases (CDs) are by far the best characterized enzymes that provide sulfur for incorporation into biomolecules. They catalyze the PLP-dependent β-elimination of sulfur from cysteine, yielding alanine and an enzyme-bound persulfide (Figure II.1). Sequence similarities divide the cysteine desulfurases into two distinct groups[27]. Group I enzymes, which include IscS and NifS[27], are usually essential for survival of the organism, and are generally thought to be responsible for the majority of iron-sulfur cluster biosynthesis[6,7,28,29]. On the other hand, the Group II enzymes (known as SufS) are typically nonessential and have lower cysteine desulfurase activity in comparison with the Group I enzymes, an observation that led to their initial designation as selenocysteine lyases (because their activity towards cysteine is lower than toward selenocysteine). The low observed activity of these enzymes could, however, be related to the role of the *suf* operon in repair of oxidative damage to iron-sulfur clusters[30-33]. In fact, SufS – and not the otherwise encoded Group I enzyme – has proven essential in several organisms, including *Synechocystis*[1], *Bacillus subtilis*[34] and *Mycobacterium smegmatis*[35], suggesting that these organisms rely on the *suf* machinery for iron-sulfur cluster biosynthesis rather than the *isc* machinery. To date, structures of several cysteine desulfurases have been published: NifS from *T. maritima*[36], IscS from *E. coli* (both group I enzymes)[37] and SufS from *E. coli* (a group II enzyme)[38-40]. The structure of C-DES from *Synechocystis*, a PLP-dependent lyase that catalyzes the same reaction by a slightly different mechanism, has also been published[41,42]. Structurally, all of the cysteine desulfurases are made up of two domains: a larger domain that binds PLP and a smaller domain that harbors the active site cysteine.

In this study, the cysteine desulfurase of interest is encoded in the *suf* operon. Appearance of the *suf* operon (in the absence of an *isc* operon) in ancient organisms hints

that it could be the precursor to the more modern cluster biosynthetic machinery, making it a good system for investigation[33]. In comparison to the Isc and Nif systems, it is thought to be the simplest of the three, and subsequent investigation into its putative role in oxidative damage repair will prove fascinating. Additionally, genetic evidence suggests that the *suf* machinery is the primary source of iron-sulfur clusters in plant chloroplasts[43-49]. As more of the enzymes involved in cluster bioassembly are biochemically and structurally characterized, it will be quite interesting to identify the features that enable the *suf* operon to operate under oxidative conditions, if that is indeed its role, and to investigate the other differences between the separate cluster biosynthetic machineries that allow such specialization.

SufS from *Synechocystis* (sySufS) is a Group II CD[27]. In addition to sySufS, two group I CDs, Slr0387 and Sll0704, are encoded in the Synechocystis sp. PCC 6803 genome[50,51], and unlike *A. vinelandii*, *H. pylori* and *S. cerevisiae*, only the group II CD is essential to survival of the cyanobacterium[6,7,28,29,52]. Careful biochemical characterization of sySufS was conducted by our collaborators B. Tirupati and J.M. Bollinger to begin to understand the detailed mechanism of iron-sulfur cluster assembly in chloroplasts[1]. Briefly, their findings, which were published with our structure[1], are as follows.

The as-isolated form of sySufS contained 0.58 equivalents of PLP and exhibits the expected absorption maxima at 425 nm for the cofactor, which could be reduced by addition of $NaBH_4$[1]. Addition of substoichiometric amounts of L-cysteine leads to decrease in the absorption at 425 nm and an increase at 342 nm, after which the spectrum slowly relaxes to the resting state. This relaxation does not occur when the active site cysteine, $C_{372}$, is mutated to alanine. An analysis of these changes has led to the conclusion that substrate binds to cause the first spectral change, and then a slow

chemical step, possibly cleavage of the substrate C-S bond, cleavage of the persulfide intermediate or release of the product alanine, results in regeneration of the first spectrum[1].

To determine which chemical step is the rate-limiting one that causes such a slow return to the resting state, the individual steps were characterized[1]. Rate of the attack of $C_{372}$ on the substrate sulfur was monitored by formation of radiolabeled enzyme with $[^{35}S]$-L-cysteine, while the persulfide cleavage step was followed via the volatilization of sulfur. These experiments confirmed that incorporation of the radiolabel into sySufS occurred on the same time scale as the slow spectral change, consistent with attack by $C_{372}$ on the substrate being the rate-limiting step. In addition to these experiments, steady-state kinetic analysis of sySufS and $C_{372}$A-sySufS showed that the mutant enzyme retained some activity in the presence of dithiothreitol and excess concentrations of L-cysteine. This activity was attributed to attack on the substrate cysteine sulfur atom by free L-cysteine, which was shown to occur in experiments with both the wild type and mutant sySufS[1].

The preliminary kinetic characterization[1] has shown that formation of the cysteinyl persulfide intermediate by sySufS occurs at a rate 1,000-fold slower than when catalyzed by a group I CD[53]. The slow rate suggests that an accessory factor, such as SufE[54,55], must be required to accelerate attack of the active site cysteine on the substrate. In the remainder of this chapter, we describe the structure of sySufS and discuss its mechanistic implications.

## II.C. Results and Discussion

### II.C.1. Overall features of PLP-bound SufS

46

Molecular replacement was used to solve the structure of sySufS, a Group II cysteine desulfurase, to 1.8 Å resolution. SufS from *Synechocystis* PCC sp. 6803 is a homodimeric protein structurally homologous to *E. coli* SufS[39] (ecSufS) (rmsd 1.1 Å over 401 Cα atoms, 50% sequence identity, all as calculated by Dali[56]), *T. maritima* NifS[36] (rmsd 2.1 Å over 353 Cα atoms, 27% sequence identity), and other PLP-dependent proteins, such as C-DES[41] (rmsd 2.3 Å over 361 Cα atoms and 26% sequence identity) from *Synechocystis*.

Pyridoxal 5'-phosphate is observed covalently linked to $Lys_{231}$ of each monomer, and a molecule of glycerol can be seen in electron density maps in each active site where substrate is expected to bind based on previous structures of ecSufS[39] and other related PLP-dependent enzymes. In addition to this, a well-ordered detergent molecule appears bound to one of the monomers at a crystal lattice contact.

### II.C.1.a. Overall Fold

Structurally, sySufS is a member of the Fold Type I family of PLP-dependent enzymes (also referred to as the α-family)[57-59]. Like the rest of those enzymes, sySufS consists of two domains: a large domain that binds PLP, and a smaller domain that harbors the active site cysteine (see Figure II.2a for the overall topology of a sySufS monomer (numbering of secondary structural elements is in agreement with that used previously[60]), and Figure II.2b for the fold of the dimer). The main fold of the large N-terminal domain is a seven-stranded β sheet with topology AGFEDBC. This sheet is surrounded on both sides by α helices (Figure II.2a). The individual strands are all parallel, with the exception of strand G, which follows the loop that contains the lysine that links to PLP in

the resting enzyme state ($K_{231}$). PLP binds at the C-terminal side of the β sheet, with its phosphate group bound at the N-terminal end of helix 5. The smaller C-terminal domain adopts a three-stranded antiparallel β sheet with α helices packed on one side and one helix (helix 17) on the "inner" side, closer to the active site. A loop preceding strand f of this domain harbors the active site cysteine $C_{372}$ and helix 17. The N-terminal 35 residues of sySufS are also part of this smaller C-terminal domain.

The sySufS dimer is formed by oligomerization at the N-terminal larger domain, with an interaction surface provided mainly by helices 5, 8 and 14 along with strands H and I. Some contacts are made with the large domain's main β sheet (Figure II.2b) and with the sySufS smaller domain. The dimerization surface stretches quite far along both monomers, allowing extensive interactions between the two molecules. Differences in the overall structure of sySufS with respect to ecSufS[38-40] are limited to short insertions and deletions such as a short insertion between strands H and I, and several minor differences in secondary structure assignment (see Figure II.3).

### II.C.1.b. Comparison to other PLP-binding protein folds

The cysteine desulfurases are members of the largest and best-characterized family of the five different PLP-dependant enzyme folds, referred to as fold type I, the α-family or the aspartate aminotransferase family[39,59,61] (Figure II.4), represented here by diaminopelargonic acid synthase[62]. These enzymes are typically homodimers (though larger oligomeric complexes are known) and contain two domains, as described above for sySufS. In this fold type, the pyridine nitrogen is typically coordinated by an aspartate residue, though other interactions are not strictly conserved, suggesting fine-tuning of each enzyme for its individual reaction. The PLP-dependant enzyme fold type I family

can be further divided into six subclasses based on the details of their N-terminal domain structure[59], with sySufS a member of the cystathionine synthase-like subclass.

Fold type II, or the tryptophan synthase β family[57-59], is characterized by a PLP-binding region formed by two β sheets (one four-stranded and one six-stranded), both surrounded by helices (Figure II.5), as observed in the tryptophan synthase structure[63]. These two sheets form two separate domains of the protein, and each member of this family typically contains a separate regulatory subunit or domain in addition to the PLP-binding domain or subunit. A cleft between the two β sheets accommodates the PLP binding site, where PLP is bound with the re side facing away from the protein and a serine ligating the pyridine nitrogen.

The third PLP-dependant enzyme fold type is the alanine racemase family[57-59]. These enzymes are also typically homodimeric, with two domains per chain – a TIM barrel domain, and a second mainly-β domain (Figure II.6)[64]. PLP binds in a cleft formed between the two domains, with its phosphate group at the N-terminus of the last TIM barrel helix (as in the fold types I and II enzymes) and its pyridine nitrogen typically bound by an arginine residue.

D-amino acid aminotransferase and branched-chain aminotransferase are members of the PLP-dependant fold type IV[57-59]. Like fold type I, these enzymes consist of two domains, with the PLP binding site formed between the two domains[65]. The N-terminal domain is smaller, with a six-stranded antiparallel β sheet, while the larger C-terminal domain folds into a mixed β sheet that wraps into a partial β barrel (Figure II.7). This second domain forms the PLP binding site, with the PLP pyridine nitrogen ligated by a glutamate residue.

Glycogen phosphorylase represents the fifth PLP-dependant enzyme fold type[59], though it uses the phosphate of PLP for proton transfer rather than using the cofactor in the typical way. Glycogen phosphorylase has three domains, with PLP binding at the C-terminal domain (Figure II.8)[66]. The phosphate group is again bound at the N-terminus of an α helix, and in this fold type, no hydrogen bonds are made to the pyridine nitrogen.

*II.C.2. Active site of SufS*

*II.C.2.a. Location and general description of the active site*

The sySufS active site is located at the dimerization interface and between the large and small domains. Residues from both monomers of the dimer make up the active site of sySufS. The majority of the active site is made up of residues from the first molecule, but several important interactions to the substrate are provided by the second molecule; in particular, residues located on the partially disordered loop connecting helices 3 and 4, and the loops before and after the β turn motif (Figure II.9). The PLP binding site is provided primarily by the larger N-terminal domain (see below), while the residues involved in catalysis, including $C_{372}$, originate from the smaller C-terminal domain or the second molecule. When bound to sySufS, the substrate is likely oriented in the active site for catalysis by $N_{180}$, $R_{367}$, $R_{387}$, $N_{58b}$ and $E_{259b}$. Nucleophilic attack is carried out by $C_{372}$, and stabilization of the resulting persulfide intermediate is made possible by $H_{129}$, $H_{370}$, and the amino group of $E_{259b}$.

Most of the active site residues are conserved between sySufS and ecSufS[39]. Differences include several conservative substitutions (such as replacement of an alanine residue that packs against the PLP with cysteine), and increased disorder in the loop between helices 3 and 4, which in ecSufS make up one side of the active site cleft. This

disorder was not surprising, as the residues in that loop have been implicated in facilitating substrate diffusion[39]; however, it may be involved with interactions between sySufS and its target protein.


*II.C.2.b. PLP Binding*

Numerous contacts in the sySufS active site, several of which are shown in Figure II.9, bind PLP tightly. Notable residues include $K_{231}$, which forms an imine linkage with the PLP aldehyde in the resting state, $D_{209}$, providing hydrogen bonds to the pyridoxal nitrogen, and $H_{128}$ and $C_{207}$, packing against the pyridine ring. Figure II.10 shows a composite omit map contoured at 1 $\sigma$ focused on the area of the PLP linkage. The bond between C4$\alpha$ of the cofactor and N$\zeta$ of $Lys_{231}$ is not coplanar with the pyridine ring of the cofactor. This nonplanarity was unanticipated, as small molecule crystallography has shown that the oxidized form of the coenzyme has a conjugated ring system which extends to the imine linkage and coplanarity has been observed in enzymes such as phosphoserine aminotransferase[67], though several PLP-dependant enzymes, including lysine 5,6-aminomutase, also exhibit this nonplanarity[60,68,69]. Because the crystals were yellow, it was assumed that the linkage was in the oxidized form, since the reduced form is colorless. However, electron density maps calculated from the data collected on a crystal treated with sodium borohydride clearly indicates that the PLP-lysine linkage has virtually the same stereochemistry in both the reduced and "oxidized" crystals. Attempts to crystallize protein that was pre-reduced with sodium borohydride were not successful.

One possible explanation of this observed nonplanarity is that the enzyme as we have crystallized it is indeed in its reduced form, and that the yellow color of the crystals arises from unbound PLP in solution adventitiously bound within the crystal. However,

the enzyme is purified with PLP bound, and no PLP was added during the purification[1] or

crystallization. Another possible explanation is photoreduction of the linkage by X-rays

during data collection, which could occur during data collection using synchrotron

radiation.

### II.C.2.c.The Active Site $C_{372}$

In the cysteine desulfurases, an active site cysteine performs a nucleophilic attack on the

sulfur atom of the substrate cysteine, resulting in $\beta$-elimination of the sulfur atom and

formation of a protein-bound persulfide. This active site cysteine residue, $C_{372}$, resides

on a short loop emanating from the smaller C-terminal domain of sySufS. The persulfide

is stabilized in sySufS by interactions with $H_{129}$, $H_{370}$ and the amino backbone atom of

$E_{259b}$ (Figure II.9). In our structure, the enzyme is in its resting state, with a water

molecule occupying the site that would presumably accommodate the persulfidic sulfur

atom. We are certain that the density observed in this site corresponds to a water

molecule, and not a sulfur atom, because the distance observed between this density and

the sulfur atom of $C_{372}$ is longer (3.4 Å) than we would expect to see for a true persulfide

(2.3 Å). In addition, refinement of the model with a water molecule in this site results in

more reasonable B-factors than are obtained when the model is refined with a persulfide

at $C_{372}$.

Attempts were also made to structurally characterize sySufS in substrate-bound

and intermediate-bound forms by cocrystallizing both the wild type and a $C_{372}A$ variant

of the enzyme with L-cysteine, L-alanine and L-serine. Only one dataset of good quality

was obtained from these experiments, and was collected from the $C_{372}A$ enzyme

cocrystallized with L-cysteine. However, the resulting electron density was unclear around the active site, so no publishable model was refined.

The enzyme under study here, sySufS, catalyzes the same reaction and has the same general fold as the Class I CDs, such as IscS from *T. maritima* (tmIscS). Because biochemical characterization showed that sySufS was inefficient in catalysis in comparison to tmIscS, a brief comparison of the structures of the two enzymes was carried out to identify any structural features that may explain the biochemical observations. Several differences within the active sites become apparent when comparing the sySufS structure to that of tmIscS[36] (Figure II.11). Most notably, the active site cysteine and the loop upon which it resides is disordered in the IscS structure, and the finger loop created by strands H and I in SufS (two short antiparallel β strands, colored yellow at top right of Figures II.9 and II.11) is missing in IscS. As one can see in Figure II.11, the two areas described above are close in space, and the presence of one could possibly make up for (with respect to the dimerization surface) the absence of the other.

### II.C.2.d. N-octanoylsucrose binding

Another point of interest in the sySufS structure is the location of the bound detergent molecule (n-octanoylsucrose). The sucrose moiety is visible at a crystal lattice contact, while the hydrocarbon chain can be seen inserted between the two domains of chain A, the last carbon positioned 11 Å from the active site cysteine (Figures II.12, II.13). As it is observed in only chain A of the two molecules per asymmetric unit, chain B was used to investigate differences caused by detergent binding. Structural rearrangements in the area surrounding the detergent are limited to movement of Phe (2 Å), and slight

(approximately 0.5 Å) adjustment of the backbone and side chains immediately adjacent to the detergent. No structural rearrangements are propagated to the active site, and it is currently unclear whether this binding has any functional relevance. However, the detergent has a mostly-hydrophobic binding pocket (Figure II.12) into which it fits quite well, and the residues involved in this binding site are fairly highly conserved (Figures II.14, II.15), suggesting that this area of the protein structure may have some importance.

## II.C.3. Implications for Catalysis

The biochemical characterization of sySufS has shown that, in comparison to the Group I cysteine desulfurases like IscS, the Group II enzymes catalyze desulfurization at a slower rate as a result of a less efficient nucleophilic attack on the substrate by $C_{372}$[1,53-55,70]. This inefficiency may arise from a shorter, more constrained active site cysteine loop in the SufS enzymes, as observed here and in structures of SufS from other organisms[38-40] (Figures II.2 and II.11). The extra flexibility conferred by the longer, disordered loop in IscS makes it physically more capable of nucleophilic attack on the substrate sulfur, while the active site cysteine is more restrained in SufS, making it less active toward the cysteine substrate[40]. This hypothesis fits with the observed activities of the enzymes: because selenium is more easily polarizable than sulfur, SufS is better capable of catalyzing the elimination reaction with selenocysteine than with cysteine[27,40]. A plausible hypothesis would therefore be that SufS requires accessory factors (i.e. SufE[54,55]) to increase its activity toward cysteine. Because of the high conservation observed at the n-octanoylsucrose binding site (Figure II.15), this site represents one possible location on SufS that an accessory factor could be expected to bind.

## II.D. Materials and Methods

### II.D.1. Protein crystallization and data collection

Protein was obtained from Ms. Bhramara Tirupati in the laboratory of J. M. Bollinger (Pennsylvania State University). Protein purification was carried out as described[1]. Standard hanging drop vapor diffusion methods were used to crystallize sySufS in a cold room at 4°C over a reservoir solution of 0.1 M Tris pH 8.0, 0.25 M MgCl$_2$, and 25% PEG 4000. Optimal crystals (Figure II.16) were grown from drops made by mixing 2μL protein stock solution (17 mg/mL sySufS in 100 mM HEPES pH 7.8 and 3% glycerol), 0.5 μL adjusted well solution (0.1 M tris pH 8.0, 0.3 M MgCl$_2$, 28% PEG 4000), 0.2 μL microseeding solution (1:10,000 dilution of crystals grown over the above reservoir solution), and 0.3 μL of a 44 mM n-octanoylsucrose solution (Hampton Research). Crystals grown from these conditions appeared in 1-2 days and had an intense yellow color, presumably due to bound PLP. Prior to data collection, crystals were transferred to reservoir solution supplemented with 20% glycerol for 30 seconds and flash cooled in liquid nitrogen.

The crystals belong to the space group P2$_1$2$_1$2$_1$ with unit cell dimensions a=88.9 Å, b=89.3 Å, and c=142.0 Å, and contain two molecules of sySufS per asymmetric unit. Data were collected to 1.8 Å resolution at ALS beamline 5.0.2 and processed using DENZO and SCALEPACK[71] (see Table II.1 for data collection statistics).

### II.D.2. Structure determination and model refinement

The structure of sySufS was determined by molecular replacement with AMoRe[72] using a polyalanine-substituted model derived from ecSufS (PDB code 1JF9)[39]. Refinement was carried out using CNS[73], protein model building was done in

XtalView[74,75], and Quanta (Molecular Simulations, Inc) was used for the construction and fitting of the n-octanoylsucrose detergent molecule. Of 420 residues per monomer, residues 7-414 were observed in electron density maps, with one chain break per molecule (between 59 and 62 of chain A, and between 59 and 63 of chain B). See Table II.2 for refinement statistics.

*II.D.3 Cocrystallization experiments*

Cocrystallization experiments were carried out in an attempt to obtain a structure of sySufS in complex with substrate or reaction intermediates. Several protein samples were used: the wild type sySufS (wt), the $C_{372}A$ mutant sySufS ($C_{372}A$), and the wild type sySufS reduced with $NaBH_4$ (red-wt). The amino acids used for cocrystallization were obtained from Hampton Research. Cocrystallization experiments were set up in different ways. First, the wt and $C_{372}A$ samples were cocrystallized with L-cysteine only, at 0.5, 1.0, 1.5, 2.5, 5.0 and 10.0 mM, all with the n-octanoylsucrose additive. Both hanging and sitting drops were set up, and drops were set up both with and without microseeding. The concentrations of all components of the previous crystallization conditions (section II.D.1) were maintained. Next, the wt and C372A were cocrystallized with L-cysteine, L-alanine and L-serine. In this case, a 3-fold excess of the amino acid was pre-incubated with the protein for 15 minutes prior to crystallization. Drops were set up maintaining the crystallization conditions described in section II.D.1, both with and without microseeding. Finally, wild-type sySufS crystals were reduced with $NaBH_4$ and cocrystallized with L-cysteine. An aqueous solution of 10 mM $NaBH_4$ in 0.24 % NaOH was added to wild type sySufS and monitored by UV-Vis spectroscopy at 335 and 425 nm to ensure that reduction of the enzyme-PLP linkage was complete. After reduction,

the buffer was exchanged to 100mM HEPES pH 7.8 and 3% glycerol, the enzyme was

pre-incubated with L-cysteine and cocrystallization experiments were set up as described

above.

One crystal (slr62) obtained by cocrystallization of C372A-sySufS with L-

cysteine was of sufficient quality to collect data. This data (Table II.3) was phased by

molecular replacement with AmoRe, yielding a model that was refined in CNS to $R/R_{free}$

= 18.45 / 22.87. See Appendix 1 for a description of SufS / SufE cocrystallization

experiments.


*II.D.4. Reduction of sySufS crystals with NaBH₄*

Crystals of sySufS obtained as described in section II.D.1 were reduced as

follows. 0.5μL of a 1M solution of NaBH₄ was added to a drop containing one single

and yellow sySufS crystal. This crystal (slr70) was cryoprotected and frozen as described

in II.D.1 and used to collect 158 frames of data, at which frame crystal decay prevented

further useful data collection. The data from slr70 extended to 1.95 Å resolution (Table

II.3). Phasing was accomplished by molecular replacement, and the model was refined in

CNS without PLP to $R/R_{free}$ = 24.39 / 26.51. Maps calculated from the resulting model

clearly contained PLP in the same orientation as the published model.


## II.E.  References

[1]     Tirupati, B., Vey, J. L., Drennan, C. L., and Bollinger, J. M., Jr., Kinetic and
        structural characterization of Slr0077/SufS, the essential cysteine desulfurase
        from Synechocystis sp. PCC 6803. *Biochemistry* **43** (38), 12210 (2004).

[2]     Jacobson, M R.; Cash, V L.; Weiss, M C.; Laird, N F.; Newton, W E.; Dean, D R.
        , Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter
        vinelandii *Molecular and General Genetics* **219** (1-2), 49 (1989).

[3]  Zheng, L. et al., Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc Natl Acad Sci U S A* **90** (7), 2754 (1993).

[4]  Zheng, L. and Dean, D. R., Catalytic formation of a nitrogenase iron-sulfur cluster. *J Biol Chem* **269** (29), 18723 (1994).

[5]  Flint, D. H., Escherichia coli contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the nifS gene of Azotobacter vinelandii and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase. *J Biol Chem* **271** (27), 16068 (1996).

[6]  Zheng, L., Cash, V. L., Flint, D. H., and Dean, D. R., Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA-fdx gene cluster from Azotobacter vinelandii. *J Biol Chem* **273** (21), 13264 (1998).

[7]  Li, J. et al., Yeast mitochondrial protein, Nfs1p, coordinately regulates iron-sulfur cluster proteins, cellular iron uptake, and iron distribution. *J Biol Chem* **274** (46), 33025 (1999).

[8]  Agar, J. N. et al., IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* **39** (27), 7856 (2000).

[9]  Krebs, C. et al., IscA, an alternate scaffold for Fe-S cluster biosynthesis. *Biochemistry* **40** (46), 14069 (2001).

[10]  Kispal, G., Csere, P., Prohl, C., and Lill, R., The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J* **18** (14), 3981 (1999).

[11]  Kurihara, T. et al., Assembly of iron-sulfur clusters mediated by cysteine desulfurases, IscS, CsdB and CSD, from Escherichia coli. *Biochim Biophys Acta* **1647** (1-2), 303 (2003).

[12]  Begley, T. P. et al., Thiamin biosynthesis in prokaryotes. *Arch Microbiol* **171** (5), 293 (1999).

[13]  Lauhon, C. T. and Kambampati, R., The iscS gene in Escherichia coli is required for the biosynthesis of 4-thiouridine, thiamin, and NAD. *J Biol Chem* **275** (26), 20096 (2000).

[14]  Leimkuhler, S. and Rajagopalan, K. V., A sulfurtransferase is required in the transfer of cysteine sulfur in the in vitro synthesis of molybdopterin from precursor Z in Escherichia coli. *J Biol Chem* **276** (25), 22024 (2001).

[15]  Kambampati, R. and Lauhon, C. T., IscS is a sulfurtransferase for the in vitro biosynthesis of 4-thiouridine in Escherichia coli tRNA. *Biochemistry* **38** (50), 16561 (1999).

[16]  Mueller, E. G., Palenchar, P. M., and Buck, C. J., The role of the cysteine residues of ThiI in the generation of 4-thiouridine in tRNA. *J Biol Chem* **276** (36), 33588 (2001).

[17]  Lauhon, C. T., Requirement for IscS in biosynthesis of all thionucleosides in Escherichia coli. *J Bacteriol* **184** (24), 6820 (2002).

[18] Kessler, D., Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes. *FEMS Microbiol Rev* **30** (6), 825 (2006).

[19] http://www.brenda-enzymes.info/.

[20] Gentry-Weeks, C. R., Keith, J. M., and Thompson, J., Toxicity of Bordetella avium beta-cystathionase toward MC3T3-E1 osteogenic cells. *J Biol Chem* **268** (10), 7298 (1993).

[21] Leibrecht, I. and Kessler, D., A novel L-cysteine/cystine C-S-lyase directing [2Fe-2S] cluster formation of Synechocystis ferredoxin. *J Biol Chem* **272** (16), 10442 (1997).

[22] Lang, T. and Kessler, D., Evidence for cysteine persulfide as reaction product of L-Cyst(e)ine C-S-lyase (C-DES) from Synechocystis. Analyses using cystine analogues and recombinant C-DES. *J Biol Chem* **274** (1), 189 (1999).

[23] Tchong, S. I., Xu, H., and White, R. H., L-cysteine desulfidase: an [4Fe-4S] enzyme isolated from Methanocaldococcus jannaschii that catalyzes the breakdown of L-cysteine into pyruvate, ammonia, and sulfide. *Biochemistry* **44** (5), 1659 (2005).

[24] Ravot, G. et al., rdlA, a new gene encoding a rhodanese-like protein in Halanaerobium congolense and other thiosulfate-reducing anaerobes. *Res Microbiol* **156** (10), 1031 (2005).

[25] Spallarossa, A. et al., The "rhodanese" fold and catalytic mechanism of 3-mercaptopyruvate sulfurtransferases: crystal structure of SseA from Escherichia coli. *J Mol Biol* **335** (2), 583 (2004).

[26] Cerletti, Paolo, Seeking a better job for an under-employed enzyme: rhodanese. *Trends in Biochemical Sciences* **11** (9), 369 (1986).

[27] Mihara, H. et al., Cysteine sulfinate desulfinase, a NIFS-like protein of Escherichia coli with selenocysteine lyase and cysteine desulfurase activities. Gene cloning, purification, and characterization of a novel pyridoxal enzyme. *J Biol Chem* **272** (36), 22417 (1997).

[28] Olson, J. W., Agar, J. N., Johnson, M. K., and Maier, R. J., Characterization of the NifU and NifS Fe-S cluster formation proteins essential for viability in Helicobacter pylori. *Biochemistry* **39** (51), 16213 (2000).

[29] Schwartz, C. J., Djaman, O., Imlay, J. A., and Kiley, P. J., The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. *Proc Natl Acad Sci U S A* **97** (16), 9009 (2000).

[30] Nachin, L. et al., SoxR-dependent response to oxidative stress and virulence of Erwinia chrysanthemi: the key role of SufC, an orphan ABC ATPase. *Mol Microbiol* **39** (4), 960 (2001).

[31] Zheng, M. et al., DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. *J Bacteriol* **183** (15), 4562 (2001).

32  Wang, T. et al., The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria. *J Bacteriol* **186** (4), 956 (2004).

33  Takahashi, Y. and Tokumoto, U., A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids. *J Biol Chem* **277** (32), 28380 (2002).

34  Kobayashi, K. et al., Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A* **100** (8), 4678 (2003).

35  Huet, G., Daffe, M., and Saves, I., Identification of the Mycobacterium tuberculosis SUF machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly: evidence for its implication in the pathogen's survival. *J Bacteriol* **187** (17), 6137 (2005).

36  Kaiser, J. T. et al., Crystal structure of a NifS-like protein from Thermotoga maritima: implications for iron sulphur cluster assembly. *J Mol Biol* **297** (2), 451 (2000).

37  Cupp-Vickery, J. R., Urbina, H., and Vickery, L. E., Crystal structure of IscS, a cysteine desulfurase from Escherichia coli. *J Mol Biol* **330** (5), 1049 (2003).

38  Fujii, T. et al., Structure of a NifS homologue: X-ray structure analysis of CsdB, an Escherichia coli counterpart of mammalian selenocysteine lyase. *Biochemistry* **39** (6), 1263 (2000).

39  Lima, C. D., Analysis of the E. coli NifS CsdB protein at 2.0 A reveals the structural basis for perselenide and persulfide intermediate formation. *J Mol Biol* **315** (5), 1199 (2002).

40  Mihara, H. et al., Structure of external aldimine of Escherichia coli CsdB, an IscS/NifS homolog: implications for its specificity toward selenocysteine. *J Biochem (Tokyo)* **131** (5), 679 (2002).

41  Clausen, T. et al., Crystal structure of the cystine C-S lyase from Synechocystis: stabilization of cysteine persulfide for FeS cluster biosynthesis. *Proc Natl Acad Sci U S A* **97** (8), 3856 (2000).

42  Kaiser, J. T. et al., Snapshots of the cystine lyase C-DES during catalysis. Studies in solution and in the crystalline state. *J Biol Chem* **278** (1), 357 (2003).

43  Ellis, K. E., Clough, B., Saldanha, J. W., and Wilson, R. J., Nifs and Sufs in malaria. *Mol Microbiol* **41** (5), 973 (2001).

44  Takahashi, Y., Mitsui, A., Fujita, Y., and Matsubara, H., Roles of ATP and NADPH in Formation of the Fe-S Cluster of Spinach Ferredoxin. *Plant Physiol* **95** (1), 104 (1991).

45  Pilon-Smits, E. A. et al., Characterization of a NifS-like chloroplast protein from Arabidopsis. Implications for its role in sulfur and selenium metabolism. *Plant Physiol* **130** (3), 1309 (2002).

46    Leon, S., Touraine, B., Briat, J. F., and Lobreaux, S., The AtNFS2 gene from Arabidopsis thaliana encodes a NifS-like plastidial cysteine desulphurase. *Biochem J* **366** (Pt 2), 557 (2002).

47    Wittpoth, Claus, Kroth-Pancic, Peter G., and Strotmann, Heinrich, Over-expression and localization of an unknown plastid encoded protein in the diatom Odontella sinensis with similarities to a subunit of ABC-transporters. *Plant Science* **114** (2), 171 (1996).

48    Moller, S. G., Kunkel, T., and Chua, N. H., A plastidic ABC protein involved in intercompartmental communication of light signaling. *Genes Dev* **15** (1), 90 (2001).

49    Kushnir, S. et al., A mutation of the mitochondrial ABC transporter Sta1 leads to dwarfism and chlorosis in the Arabidopsis mutant starik. *Plant Cell* **13** (1), 89 (2001).

50    Ikeuchi, M., [Complete genome sequence of a cyanobacterium Synechocystis sp. PCC 6803, the oxygenic photosynthetic prokaryote]. *Tanpakushitsu Kakusan Koso* **41** (16), 2579 (1996).

51    Kaneko, T. et al., Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). *DNA Res* **3** (3), 185 (1996).

52    Seidler, A., Jaschkowitz, K., and Wollenberg, M., Incorporation of iron-sulphur clusters in membrane-bound proteins. *Biochem Soc Trans* **29** (Pt 4), 418 (2001).

53    Behshad, E., Parkin, S. E., and Bollinger, J. M., Jr., Mechanism of cysteine desulfurase Slr0387 from Synechocystis sp. PCC 6803: kinetic analysis of cleavage of the persulfide intermediate by chemical reductants. *Biochemistry* **43** (38), 12220 (2004).

54    Loiseau, L. et al., Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J Biol Chem* **278** (40), 38352 (2003).

55    Ollagnier-de-Choudens, S. et al., Mechanistic studies of the SufS-SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE. *FEBS Lett* **555** (2), 263 (2003).

56    Holm, L. and Sander, C., Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233** (1), 123 (1993).

57    Jansonius, J. N., Structure, evolution and action of vitamin B6-dependent enzymes. *Curr Opin Struct Biol* **8** (6), 759 (1998).

58    Denessiouk, K. A. et al., Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes. *Proteins* **35** (2), 250 (1999).

59    Schneider, G., Kack, H., and Lindqvist, Y., The manifold of vitamin B6 dependent enzymes. *Structure* **8** (1), R1 (2000).

60   Weyand, M. and Schlichting, I., Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. *Biochemistry* **38** (50), 16469 (1999).

61   Mehta, P. K. and Christen, P., The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv Enzymol Relat Areas Mol Biol* **74**, 129 (2000).

62   Kack, H. et al., Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes. *J Mol Biol* **291** (4), 857 (1999).

63   Rhee, S. et al., Exchange of K+ or Cs+ for Na+ induces local and long-range changes in the three-dimensional structure of the tryptophan synthase alpha2beta2 complex. *Biochemistry* **35** (13), 4211 (1996).

64   Shaw, J. P., Petsko, G. A., and Ringe, D., Determination of the structure of alanine racemase from Bacillus stearothermophilus at 1.9-A resolution. *Biochemistry* **36** (6), 1329 (1997).

65   Sugio, S. et al., Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity. *Biochemistry* **34** (30), 9661 (1995).

66   Acharya, K.R., Stuart, D.I.,Varvill, K.M.,Johnson, L.N. , in *Glycogen Phosphorylase B: Description of the Protein Structure* (World Scientific, 1991), pp. 1.

67   Hester, G. et al., Crystal structure of phosphoserine aminotransferase from Escherichia coli at 2.3 A resolution: comparison of the unligated enzyme and a complex with alpha-methyl-l-glutamate. *J Mol Biol* **286** (3), 829 (1999).

68   Eswaramoorthy, S. et al., Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta Crystallogr D Biol Crystallogr* **59** (Pt 1), 127 (2003).

69   Berkovitch, F. et al., A locking mechanism preventing radical damage in the absence of substrate, as revealed by the x-ray structure of lysine 5,6-aminomutase. *Proc Natl Acad Sci U S A* **101** (45), 15870 (2004).

70   Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N., Kinetic and mutational studies of three NifS homologs from Escherichia coli: mechanistic difference between L-cysteine desulfurase and L-selenocysteine lyase reactions. *J Biochem (Tokyo)* **127** (4), 559 (2000).

71   Otwinowski, Z., and Minor, W., in *Methods in Enzymology*, edited by C. W. Carter Jr., and Sweet, R. M. (Academic Press, San Diego, CA, 1997), Vol. 276, pp. 307.

72   Navaza, J., AMoRe: an automated package for molecular replacement. *Acta Crystallographica Section A* **50** (2), 157 (1994).

73   Brunger, Axel T. et al., Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica Section D* **54** (5), 905 (1998).

[74]  McRee, D.E., *Practical Protein Crystallography*. (Academic Press, San Diego, CA, 1993).

[75]  McRee, D. E., XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125** (2-3), 156 (1999).

[76]  Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F., ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15** (4), 305 (1999).

## II.F. Tables and Figures

### Table II.1: Data Collection Statistics for wild-type sySufS

[a] $R_{sym} = (\Sigma_{hkl} \Sigma_i I_i(hkl) - <I(hkl)>)/ \Sigma_{hkl} \Sigma_i I_i(hkl)$ for $n$ independent reflections and $i$ observations of a given reflection. $<I(hkl)>$ = average intensity of the $i$th observation. [b]Numbers for the highest resolution shell are shown in parentheses.

| | |
|---|---|
| wavelength (Å) | 1.100 |
| resolution (Å) | 50-1.80 |
| $R_{sym}$ (%)[ab] | 6.2 (32.4) |
| total observations | 979 036 |
| unique reflections | 99 625 |
| avg redundancy | 4 |
| $<I/\sigma>$[b] | 18.1 (4.5) |
| completeness (%)[b] | 94.4 (82.2) |

### Table II.2: Refinement Statistics for wild-type SufS

[a] $R_{cryst} = \Sigma_h F_o(h) - F_c(h) /\Sigma_h |F_o(h)|$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively. $R_{free}$ is calculated using the same equation with a test set of reflections that are not used during refinement.[b] Chain A/chain B.

| | |
|---|---|
| resolution (Å) | 50-1.80 |
| reflections (working/test) | 85,453/9531 |
| $R_{cryst}/R_{free}$ (%)[a] | 19.4/21.8 |
| residues | 811 |
| total protein atoms | 6213 |
| total ligand atoms | 81 |
| total solvent atoms | 538 |
| disordered side chains[b] | 13/9 |
| rms deviation | |
| bonds (Å) | 0.0098 |
| angles (deg) | 1.35 |
| Ramachandran analysis | |
| most-favored | 655 (91.1%) |
| allowed | 60 (8.6%) |
| generously allowed | 2 (0.3%) |
| disallowed | 0 |

**Table II.3:  Data processing statistics from crystals slr62 and slr70.**

| Crystal | slr62 | slr70 |
|---|---|---|
| Resolution (Å) | 2.6 | 1.95 |
| Rsymm | 8.5 (35.7) | 4.4 (34.1) |
| $I/\sigma$ | 10.1 | 25.8 |
| Redundancy | 5.0 | 3.0 |
| Completeness (%) | 98.2 (96.0) | 87.2 (85.2) |

**Figure II.1: The reaction catalyzed by the cysteine desulfurases** (figure was obtained from J. Martin Bollinger, Pennsylvania State University).
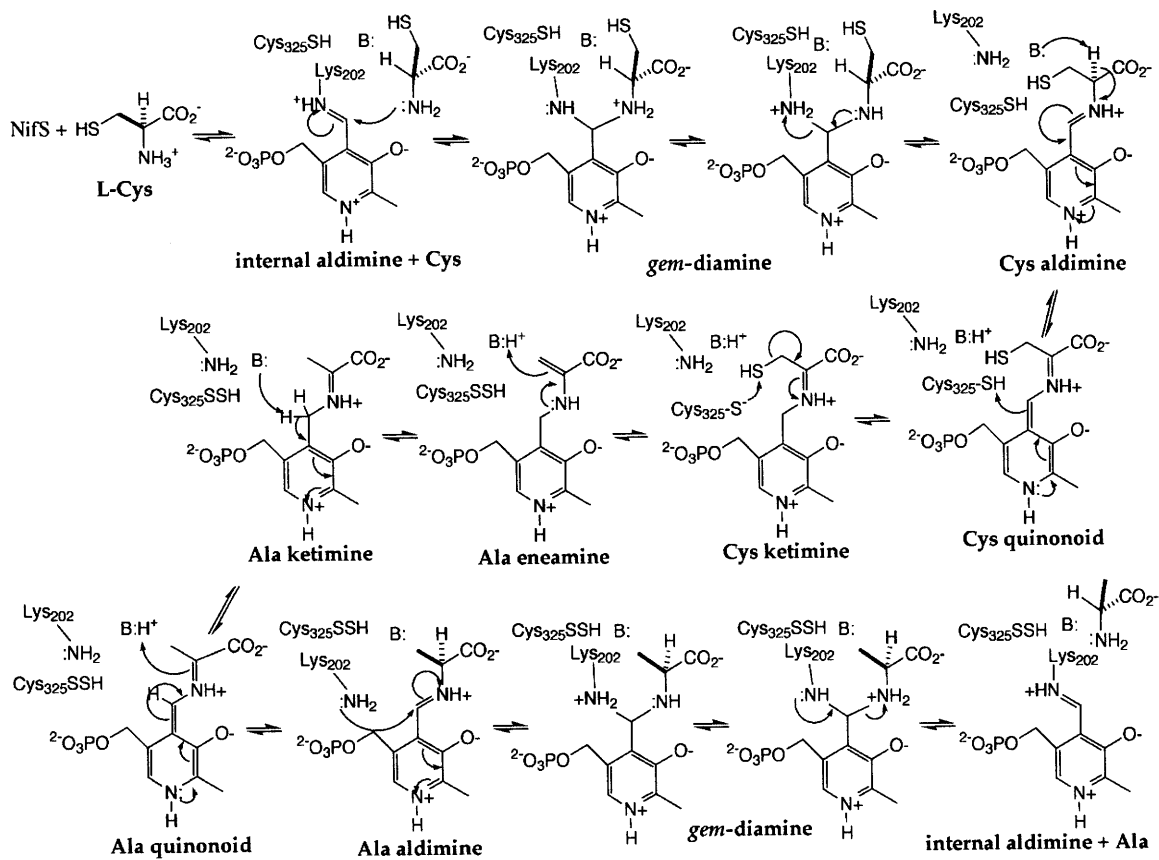
**Figure II.2: Overall fold of sySufS.** (a) Stereoview of the crystal structure of sySufS monomer with secondary structural elements labeled as in reference 60. Strands in the large domain are in capital letters, strands in the small domain are in lowercase letters, and helices are numbered. PLP, glycerol (labeled "glyc") and n-octanoylsucrose (labeled "oct") are depicted in spacefill, and active site cysteine $C_{372}$ is depicted in sticks. Several secondary structural elements in SufS differ from those in *E. coli* SufS: helices 2, 7 and 13 of *E. coli* SufS are absent in sySufS, and strand d of the *E. coli* enzyme is missing in SufS, replaced spatially by strand a′. The chain break between helices 3 and 4 is marked with two red asterisks. (b) Stereoview of the SufS dimer with one chain colored blue and the other colored yellow. PLP, glycerol, active site cysteine $C_{372}$, and n-octanoyl sucrose are depicted as in *a*, and the chain break is again denoted with two red asterisks.
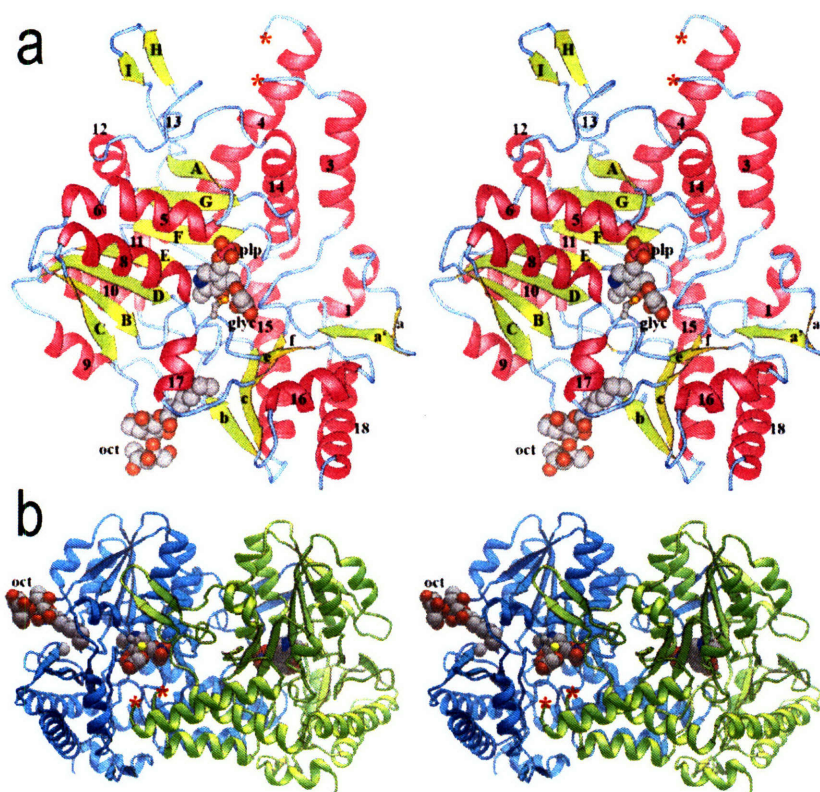
**Figure II.3: Superposition of sySufS and ecSufS in stereoview.** The cartoon representations of both enzymes are shown, with sySufS in teal and ecSufS in yellow. PLP from both enzymes is shown in spheres colored as follows: grey, carbon; blue, nitrogen; red, oxygen; orange, phosphorus.
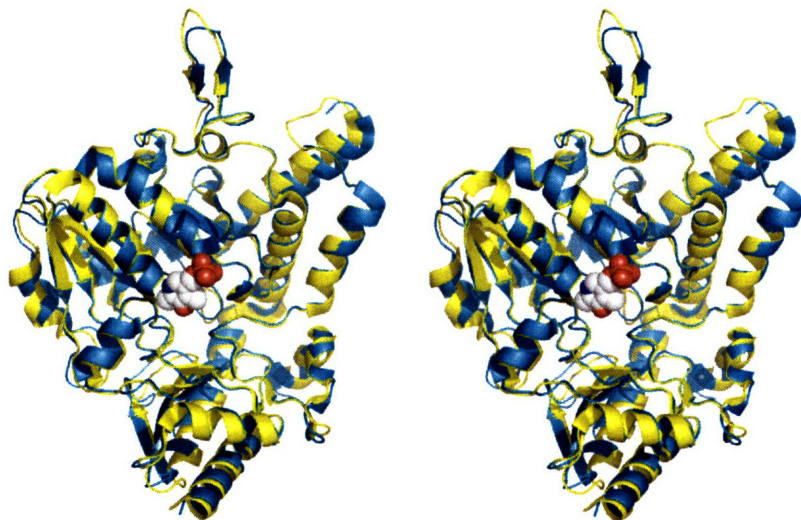
**Figure II.4: PLP-dependent enzyme fold type I**: Diamino-pelargonic acid synthase, PDB code 1QJ5. The enzyme is shown in cartoon representation with PLP shown in sticks. Strands are colored yellow, helices blue, and PLP carbons green.
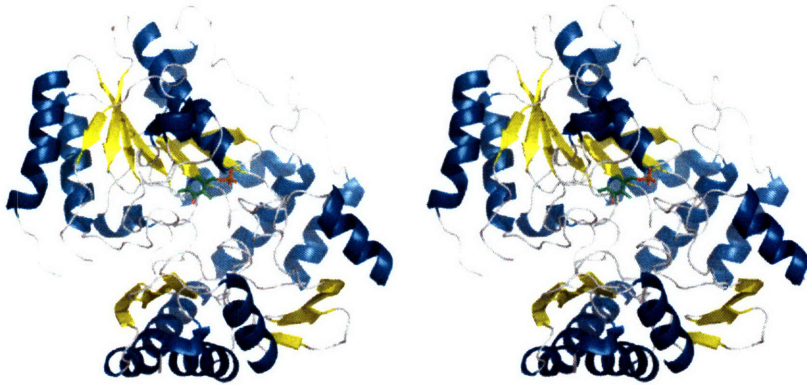


**Figure II.5: PLP-dependent enzyme fold type II**: Tryptophan synthase, PDB code 1UBS. Colors are as in Figure II.4. The regulatory subunit is colored grey.
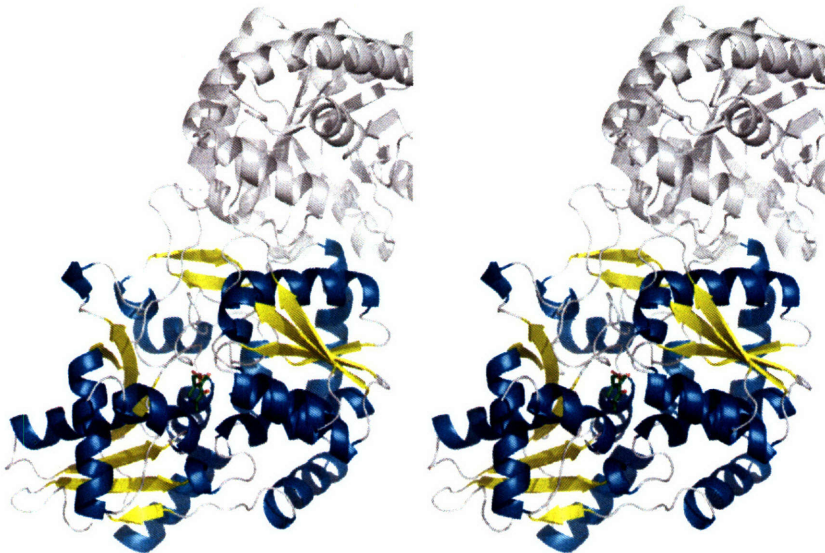
**Figure II.6:  PLP-dependent enzyme fold type III**: Alanine racemace, PDB code 1SFT.  Colors are as in Figure II.4.
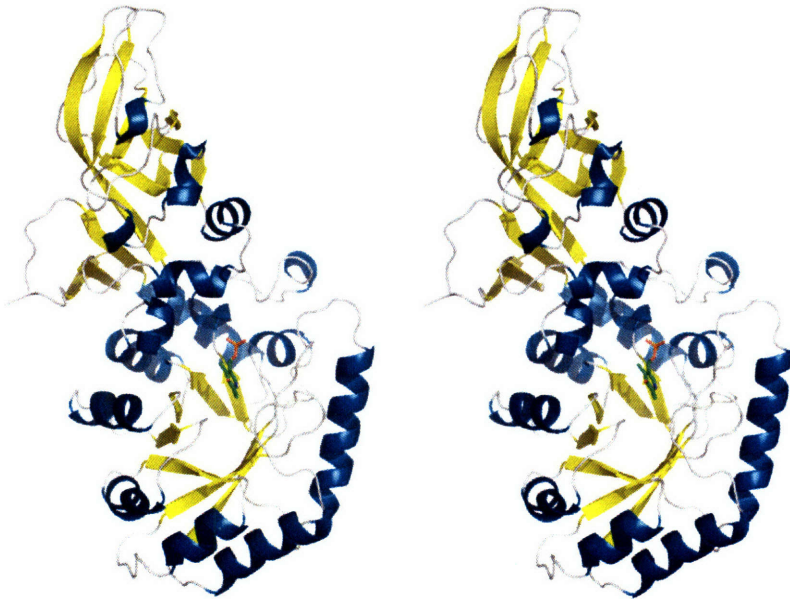


**Figure II.7:  PLP-dependent enzyme fold type IV**: D-alanine aminotransferase, PDB code 1DAA.  Colors are as in Figure II.4.
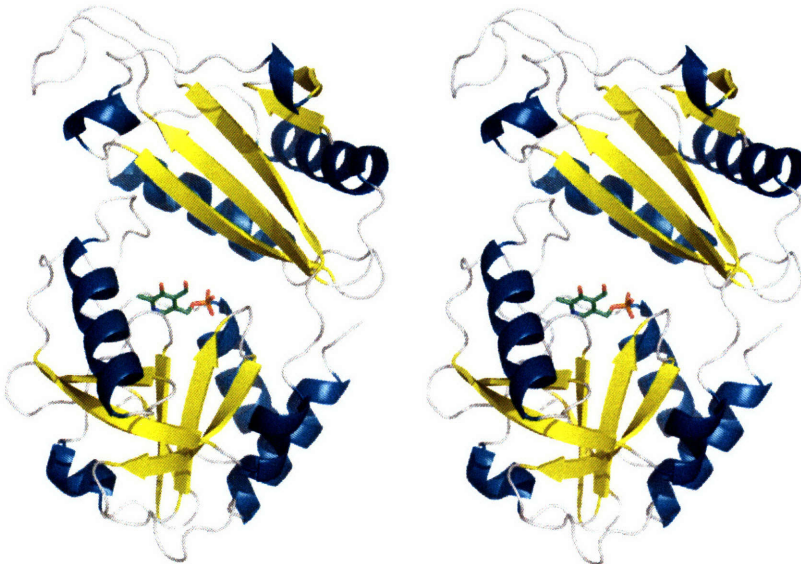
**Figure II.8:  PLP-dependent enzyme fold type V**: Glycogen phosphorylase, PDB code 1GPB. Colors are as in Figure II.4.  For clarity, the N-terminal domains are colored grey.
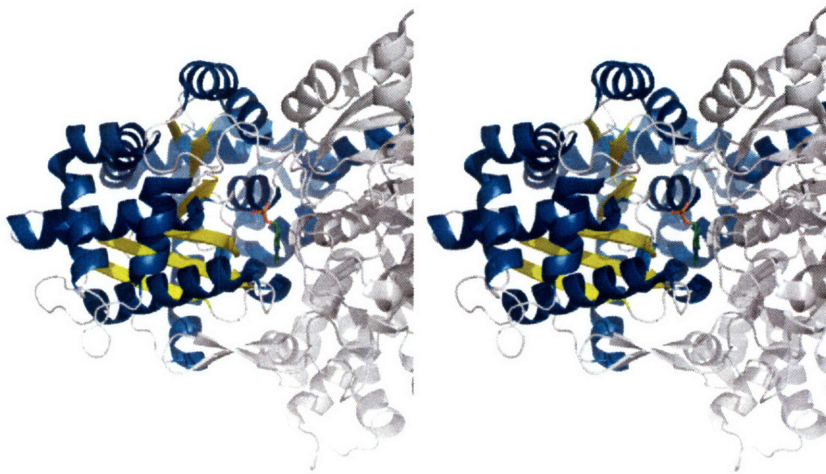
**Figure II.9:  Stereoview of the sySufS active site**.  The carbons of the PLP are colored green, and sidechains near the active site are labeled by one-letter abbreviation and number (carbons colored grey, oxygen red, nitrogen blue, sulfur yellow, phosphorus magenta).  The two monomers are colored as in Figure II.2b, and sidechains from the yellow chain are labeled with one-letter abbreviation, number, and a lowercase b.  Glycerol bound at the active site is shown in thick sticks.  Two red asterisks indicate the chain break between helices 3 and 4.
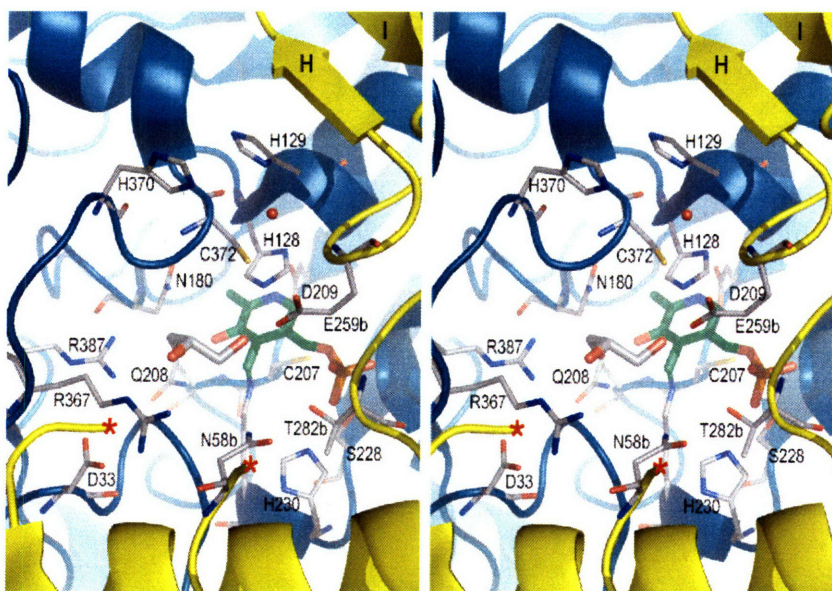


**Figure II.10:  PLP-K$_{231}$ imine linkage**.  Composite omit electron density map contoured at $1\sigma$ is shown around the covalently bound PLP cofactor and glycerol in the active site. Atoms are colored as follows:  carbon, grey; oxygen, red; nitrogen, blue.  C$\alpha$ traces are shown for both molecules in the dimer, with chains colored as in Figure II.2b.
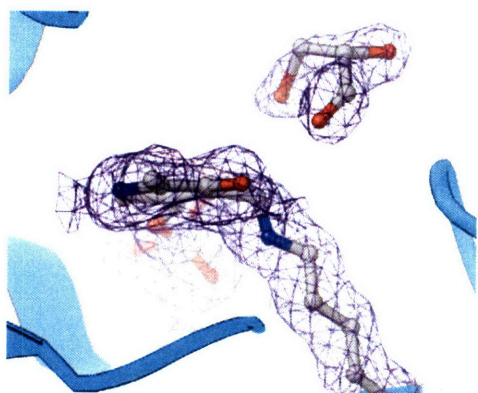
**Figure II.11: Superposition of sySufS (yellow) and IscS from *T. maritima* (purple).** $C_{372}$, PLP, and glycerol are depicted in ball-and-stick, and are colored as follows: grey, carbon; red, oxygen; blue, nitrogen; yellow, sulfur. In this figure, the red squares denote where traceable density ends as a result of the disordered IscS active site loop (see text). β strands H and I are labeled.
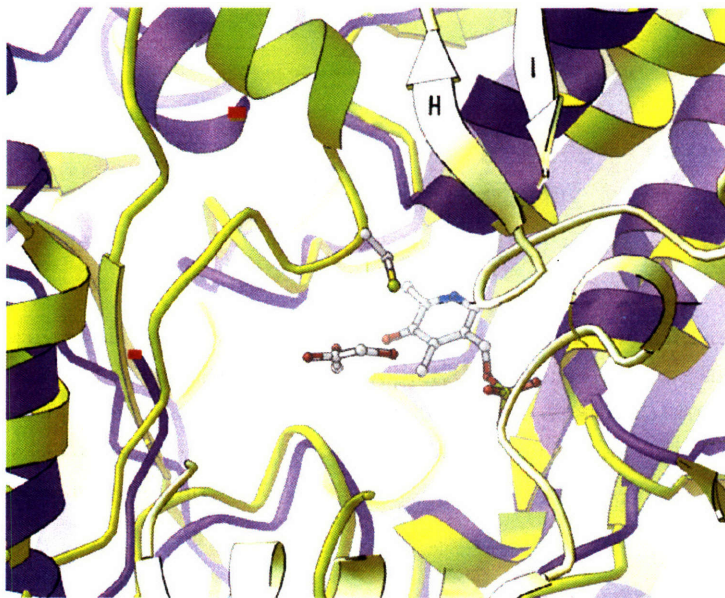


**Figure II.12: The detergent binding site.** This figure shows the electron density surrounding the detergent binding site, with a $2F_o - F_c$ omit map contoured at 1 σ in dark blue, and an $F_o - F_c$ omit map contoured at 1 σ shown in teal. The maps were calculated in CNS and omit the n-octanoylsucrose detergent only.
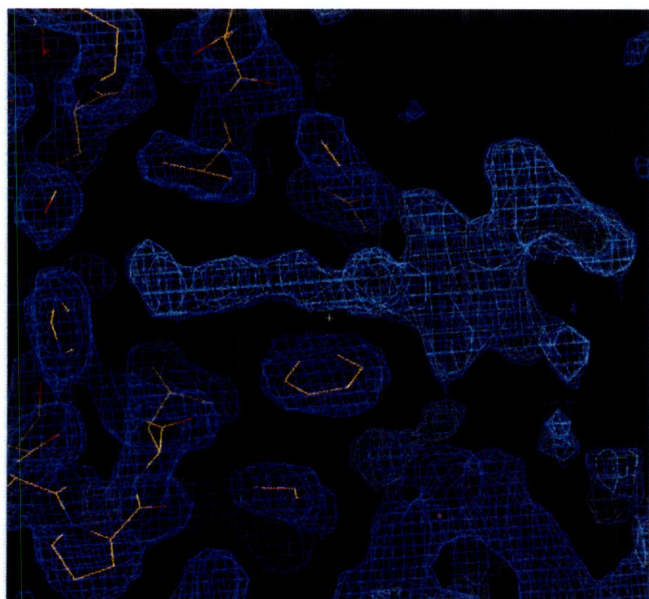
**Figure II.13: The detergent binding site.** This stereoview shows the location of the detergent binding site with respect to the N-terminal and C-terminal domains and the proximity of the n-octanoylsucrose carbon chain to the active site cysteine. sySufS is shown in the same orientation and with the same colors as in Figure II.2, with the exception that PLP, glycerol and n-octanoylsucrose are shown in stick representation, and the PLP carbons are shown in green. The active site cysteine loop is highlighted in teal, as are the carbons of $C_{372}$, which is shown in stick representation.
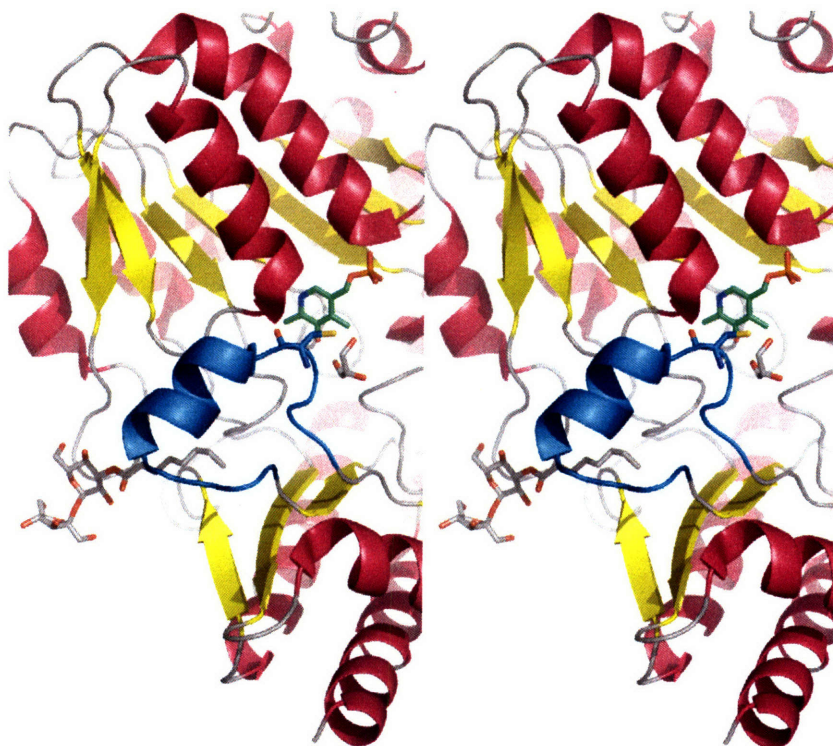
**Figure II.14: Empty detergent binding pocket**. ESPript[76] was used to calculate the sequence conservation of the sySufS residues surrounding the binding site of the n-octanoylsucrose detergent molecule. The sySufS surface near this binding site in the chain without bound detergent is displayed and colored by sequence conservation in rainbow colors, with red being 100% conserved and blue representing 0% conservation.
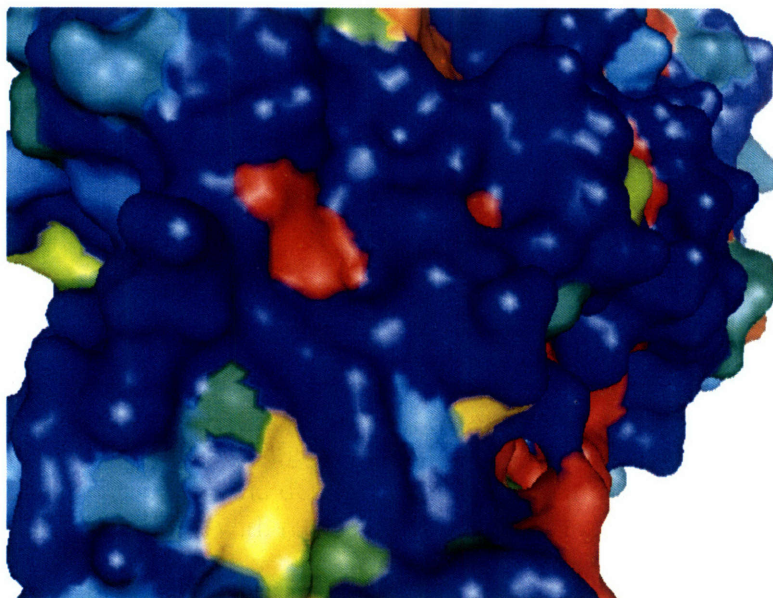


**Figure II.15: Detergent binding pocket**. The sySufS surface of the chain with bound n-octanoylsucrose is displayed and shown in the same orientation as shown in Figure II.14. The detergent molecule is shown as sticks, colored as follows: grey, carbon; blue, nitrogen; red, oxygen.
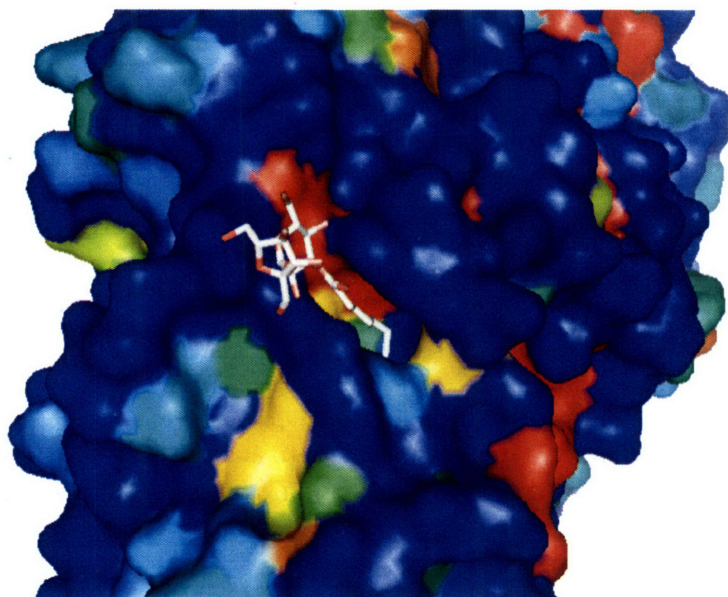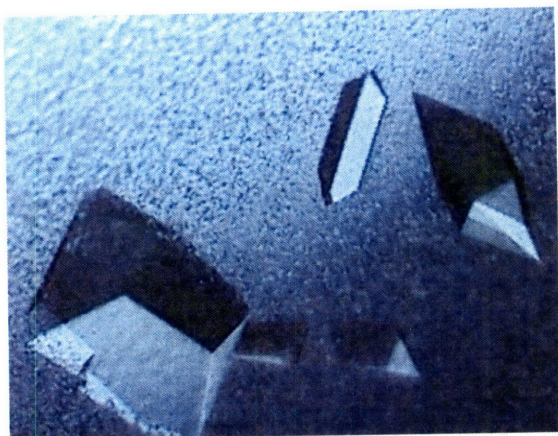
**Figure II.16: Photograph of sySufS crystals.**

# Chapter III: Structural Insights into Glycyl Radical Formation by Pyruvate formate-lyase Activating Enzyme

## III.A. Summary

Pyruvate formate-lyase activating enzyme is responsible for the generation of a catalytically essential glycyl radical on $G_{734}$ of pyruvate formate-lyase. The activating enzyme, a member of the AdoMet radical superfamily, generates this radical via the direct, stereospecific abstraction of a hydrogen atom from pyruvate formate-lyase. Insight into the interactions between the activase and the loop containing $G_{734}$ of pyruvate formate-lyase is provided by the structures reported here of the substrate-free and substrate-bound forms of the activase. These first structures of a glycyl radical-forming enzyme reveal how the glycine loop is positioned in the active site of the activase and the conformational change the activase undergoes to bind substrate.

Note: This chapter has been submitted in manuscript form for publication.

## III.B. Introduction

Pyruvate formate-lyase activating enzyme (PflAE) is a member of the "AdoMet radical" or "Radical SAM" superfamily, which is characterized by both the presence of a conserved $CX_3CX_2C$ sequence motif that coordinates an essential 4Fe-4S cluster and by the use of S-adenosyl-L-methionine (AdoMet or SAM) for 5'-deoxyadenosyl radical (5'-dA•) generation[1,2,3]. Cleavage of AdoMet to form 5'-dA• requires reduction of the 4Fe-4S cluster with one electron, which is provided in *Escherichia coli* by flavodoxin[4]. AdoMet radical enzymes act on a wide variety of biomolecules in numerous pathways. For example, biotin synthase (BioB)[5], coproporphyrinogen III oxidase (HemN)[6] and molybdenum cofactor biosynthetic enzyme A (MoaA)[7] are involved in vitamin/cofactor biosynthesis; lysine 2,3-aminomutase (LAM) facilitates the fermentation of lysine[8]; spore photoproduct lyase repairs UV-induced DNA damage[9]; and class III ribonucleotide reductase (aRNR) activase, like PflAE, catalyzes the formation of glycyl radicals[10]. The structural basis by which AdoMet radical enzymes are able to react with such a diverse group of substrates, including dethiobiotin, DNA and proteins is a key question in the field. The structure of PflAE from *E. coli*, described here, provides a first view of a member of the AdoMet radical superfamily for which the substrate is an enzyme, and thus the first activase for a glycyl radical enzyme (GRE).

Because of their oxygen-sensitive nature, GREs[11-14] are proposed to have evolved prior to the appearance of oxygen in the atmosphere, and some speculate that aRNR may be a truly ancient enzyme[15], as the original RNR was likely involved in the conversion of RNA- to DNA-based life[16]. Unlike RNR, which converts ribonucleotides to deoxyribonucleotides, Pfl is a central metabolic enzyme that converts pyruvate and coenzyme A (CoA) to acetyl-CoA and formate[17]. Found in both prokaryotes and

eukaryotes, Pfl provides the sole source of acetyl-CoA for the Krebs cycle under fermentative conditions. X-ray analysis reveals a common fold for GRE family members as a ten-stranded $\beta$-barrel, with two sets of five-stranded sheets running anti-parallel to each other and a buried glycine residue at the tip of the second of two $\beta$ finger motifs[18-23]. In one of the more fascinating activation reactions in biology, PflAE (a monomer of 28 kDa) stereospecifically abstracts a hydrogen atom from residue $G_{734}$ of Pfl (homodimer of 170 kDa), generating the catalytically relevant glycyl radical species[24]. Because $G_{734}$ is buried 8 Å from the protein surface[18], the direct abstraction of a hydrogen atom from this residue is difficult to conceptualize. Once established, the glycyl radical is surprisingly stable[25], catalyzing multiple turnovers via a putative active site thiyl radical. Though stable under anaerobic conditions, the radical-harboring glycine is susceptible to destruction by oxygen exposure, which results in irreversible cleavage of the polypeptide and inactivation of the enzyme[26]. Interestingly, some organisms have developed a mechanism to restore Pfl activity after oxygen exposure. A short protein with homology to the last 59 residues of Pfl, including $G_{734}$, is activated by PflAE and associates with cleaved Pfl to yield a working enzyme complex[27]. This short protein, called YfiD in *E. coli*, is thought to act as a "spare part" to repair oxygen-cleaved Pfl with minimal energy expenditure. Thus both Pfl and YfiD are natural substrates for PflAE in *E. coli*. *In vitro*, peptides of as few as seven residues containing the sequence surrounding $G_{734}$ can also act as substrates[24]. In order to understand how the AdoMet radical activases catalyze formation of glycyl radicals, we solved the structure of PflAE in both the substrate-free (AE) and substrate-bound (pept-AE) forms. Our results support the proposal that a small domain of Pfl harboring $G_{734}$ undergoes a

conformational change, thereby positioning $G_{734}$ in close proximity to the AdoMet bound at the active site of PflAE.

## III.C. Results

### III.C.1. PflAE Overall Fold

PflAE is a monomeric, single-domain enzyme, and with 245 amino acids, it is the smallest AdoMet radical enzyme of known structure (see Methods, Tables II.1, II.2 and Figures III.1 and III.2). Both the 2.25 Å resolution AE and the 2.8 Å resolution pept-AE structures show a partial TIM barrel fold with a 4Fe-4S cluster ligated by the canonical $CX_3CX_2C$ motif. Although the AE crystals were grown in the presence of AdoMet, this co-substrate is observed only in the pept-AE structure, where it ligates the 4Fe-4S cluster and packs against the 7-mer RVSGYAV peptide substrate (Figures III.1, III.2). PflAE consists of the $(\beta/\alpha)_6$ "AdoMet radical " core and, unlike the other AdoMet radical enzymes, few secondary structural elements outside of the core fold are present to "close off" the opening of the partial barrel (Figure III.3). Only 65 amino acids of PflAE are not part of the core region, including $\beta1'$, loop A, loop C, and $\alpha6$ (Figure III.3). The C-terminal "top" of the partial barrel is covered by two long loops located after strands $\beta1$ (loop B) and $\beta6$ (loop C). The N-terminal "bottom" and the lateral opening of the barrel appear highly solvent-exposed.

### III.C.2. AdoMet Radical Core

The AdoMet radical core fold of PflAE consists of a "splayed" sheet similar to those observed in structures of other AdoMet radical family members[5–8] (see Table III.3 for RMSDs). The active sites of these enzymes are located within the lateral opening and

are made up mainly of residues originating from the C-terminal ends of each $\beta$ strand within the core's sheet (Figure III.3). The 4Fe-4S cluster of PflAE is located at the C-terminal end of the partial barrel (Figure III.3), with the unique iron coordinated by AdoMet through the amino nitrogen (2.2 Å) and carboxyl oxygen (2.2 Å). The distances observed between AdoMet and the cluster (Figure III.4a) are in agreement with the spectroscopic data observed with this enzyme[28-30] and are similar to those seen in the other AdoMet radical structures[5-8]. These close interactions between AdoMet and the cluster are consistent with the theory that inner-sphere electron transfer can occur between the cluster of PflAE and AdoMet to generate 5'-dA•[29]. Several residues that form the AdoMet binding site are part of motifs conserved amongst all of the AdoMet radical enzymes, such as the GGE motif ($G_{77}$, $G_{78}$, and $E_{79}$ in PflAE) and the GxIxGxxE motif (PflAE residues $V_{168}$, $V_{170}$, $G_{172}$ and $E_{175}$)[31]. As discussed previously[31], these two motifs bind the AdoMet methionine directly and serve to stabilize the binding site for the adenine moiety, respectively (Figure III.4b). Additional interactions with AdoMet (Figure III.4b) are similar to those observed in other AdoMet radical enzymes, although the identity of the residues is often not conserved[5-8], a difference due in part to the large number of contacts made by backbone atoms. One interesting variation is the presence of two histidines, $H_{37}$ and $H_{202}$, which pack against the AdoMet adenine ring. As discussed below, these residues are likely important for the ability of PflAE to bind a protein substrate. Finally, $D_{104}$ and $R_{166}$ are the two protein residues that are located closest to C5' of AdoMet in this structure, but neither makes an exceedingly close hydrogen bonding interaction with either AdoMet or the peptide substrate (Figure III.4). These two residues are completely conserved in PflAE sequences. Taken together, the high conservation and suggestive positioning observed leads to the theory that these two

residues are in some way involved in catalysis through control of the 5′-dA• intermediate, orientation of the substrate or another less obvious role. This remains to be shown by mutagenesis studies.

### III.C.3. Peptide Binding

The peptide (RVSGYAV) binds in a bent conformation (Figure III.4c, III.4d) across the lateral opening of the partial barrel, interacting primarily with residues from loops A, B and C (Figure III.3). Six of the seven amino acids in the peptide are visible in the structure, with the first amino acid, $R_{731}$, disordered due to a nearby crystal lattice contact. Binding of the peptide seals the active site from solvent, burying the cluster and AdoMet and providing a protected environment for radical generation (Figures III.5, III.6). The binding of peptide also appears to stabilize AdoMet binding to PflAE, since no discernible AdoMet density is present in the substrate-free structure, despite the fact that AdoMet was present during crystallization (Figure III.1). Although previous spectroscopic studies[28, 29, 30, 32] have shown that AdoMet binds to PflAE in the absence of substrate, and our data do not explicitly contradict these results, the lack of sufficiently substantial density for AdoMet in its binding site in the substrate-free form (Figure III.7) suggests that substrate serves to stabilize AdoMet in its binding site. This possibility is attractive from a biological perspective because it provides a potential mechanism to prevent uncoupling between AdoMet cleavage and glycyl radical generation: only in the presence of substrate would AdoMet favor a stable catalytic conformation. Though the AdoMet is buried in the pept-AE structure, a large surface area including the peptide itself and the N-terminal sides of some β strands remains solvent exposed. Presumably, these surfaces will be buried in the full Pfl-PflAE complex.

The peptide substrate in the pept-AE structure corresponds to the seven-residue

consensus sequence around $G_{734}$ of Pfl: RVSGYAV ($K_m$ = 0.22 mM, $V_{max}$ = 11

nmol/min·mg, compared to 1.4 µM and 54 nmol/min·mg for Pfl)[24]. This motif is highly

conserved in Pfl. However, there is poor consensus at this site between all glycyl radical

containing enzymes, with only the residues corresponding to $R_{731}$ and $G_{734}$ being

completely conserved (Figure III.8). Specific residues on PflAE required for or involved

in the activation reaction have not been examined in detail, although it is known that the

enzyme harbors no activity towards a peptide corresponding to the similar Class III RNR

consensus sequence, RVCGYLG, which substitutes a Cys for $S_{733}$, a Leu for $A_{736}$ and a

Gly for $V_{737}$[24].

Despite some observed specificity towards the Pfl glycine loop sequence[24],

contacts are made mainly to peptide substrate backbone atoms (Figure III.4c). In

contrast, all of the contacts to the peptide are made by side chains of PflAE, each of

which are highly conserved and include a DGxGxR motif located on loop A (Figure III.3,

III.4c). The key interactions made by PflAE to bind peptide serve three purposes: (1)

orienting the glycine in the active site, (2) controlling the overall peptide conformation

and (3) imparting selectivity. Three specific residues ensure the proper orientation of

G734 with respect to AdoMet in the active site: $D_{16}$ of the DGxGxR motif and fully

conserved N38 hydrogen bond to the glycine amino and carboxyl groups, while $H_{37}$

reaches across the AdoMet adenine ring to help orient the residue (Figure III.4c). These

specific interactions ideally position $G_{734}$ for hydrogen atom abstraction, with an AdoMet

C5' to $G_{734}$ Cα distance of 4.1 Å (Figure III.4c), similar to those observed in BioB (3.7

Å)[5] and KAM (3.8 Å)[8]. The peptide conformation is dictated by the van der Waals

interactions with loop A and hydrogen bonding interactions provided by $K_{208}$ and $N_{38}$ to

$A_{736}$. Finally, selectivity appears to arise from interactions with $F_{25}$ at residue $A_{736}$ ($F_{25}$ should prevent binding of a larger amino acid) and with $L_{204}$ and $H_{207}$ at $Y_{735}$ (through packing and hydrogen bonding interactions) (Figure III.4c). Because the other interactions observed in this structure are to peptide backbone atoms, specificity at these other positions is likely governed by the effect of side chain identity on the conformation of the peptide backbone itself. Unfortunately, crystal packing prevents modeling of $R_{731}$, leaving its role in the activation reaction and the reason for its conservation unclear, though others have proposed that it is involved in the stabilization of the glycine loop structure in the native state of a GRE[19]. The pept-AE interactions observed here are not electrostatic in nature, consistent with the observation that the activation reaction is not affected by increasing the salt concentration of the activation reaction buffer[33].

Comparison of the AE and pept-AE structures (RMSD = 0.774 Å) allows us to distinguish any structural rearrangements associated with AdoMet and peptide binding (Figures III.9 – III.11). The major conformational change upon binding of the peptide occurs in loop A, which harbors the activase-specific DGxGxR motif ($D_{16}G_{17}xG_{19}xR_{21}$ in PflAE) (Figure III.9a). As the loop swings up towards the active site to form contacts with the peptide, the largest movement occurs in residue $D_{16}$, which is displaced ~10Å at the Cβ atom (Figure III.9b). $R_{21}$ of the DGxGxR motif anchors the beginning of strand β1 and may stabilize the β sheet during this motion. This movement is likely essential to the activation reaction, in enabling a structural change in Pfl and/or orienting the glycine loop in the active site. Several other side chain rearrangements or reordering occur and appear important in forming the AdoMet binding site ($H_{37}$, $L_{199}$, $H_{202}$) and binding peptide ($N_{38}$, $H_{207}$, $K_{208}$).

*III.C.4. Docking Studies*

Areas of high conservation on the PflAE surface that could be involved in protein-protein interactions, such as with Pfl or flavodoxin, were identified (see Methods). These include: (1) the active site, (2) the N-terminal side of the β sheet, and (3) behind loop B (Figure III.12a). The preservation observed at the active site (region 1) was expected and is more extensive than just the area involved in binding the peptide. Regions 1 and 2 are adjacent to one another, suggesting that the two areas may together form the PFL binding surface. The location of region 3 proximal to the 4Fe-4S cluster is ideal for interactions with flavodoxin.

A theoretical model of the mode of interaction between Pfl and PflAE was generated by manual and computational docking studies (see Methods and Figures III.13 and III.14). The best docking model was obtained using the pept-AE structure and a portion of Pfl with high homology to YfiD (the small protein capable of undergoing activation by PflAE)[27]. This region of Pfl will be referred to as a "radical domain" (RD) and corresponds to Pfl residues 712 – 759 (Figure III.12b, III.12c). In the docking model, the glycine loop of RD points into the active site of AE, positioned similarly to the peptide observed bound in the experimental maps (Figure III.12d – f, III.15), though increased conformational flexibility allows the peptide to bind in a more extended conformation towards its C-terminal end. This difference in conformations could also be due in part to the fact that RD was treated as a rigid body in the docking studies and not allowed to adjust to its new binding environment. Despite this rigid body treatment, the distance predicted between AdoMet and RD $G_{734}$ Cα (4.6 Å) in the docking model agrees reasonably well with that observed between the AdoMet and $G_{734}$ Cα of the bound peptide (4.1 Å). The interactions between $H_{37}$, $N_{38}$ or $L_{204}$ and RD are similar to those

observed in the pept-AE structure, though the bidentate interaction of $N_{38}$ and the contacts between $F_{25}$, $K_{208}$ and the peptide are lost. Residues from PflAE and RD in the docking model are also poised to make several additional interactions (Figure III.12f), including RD residue $R_{731}$ with PflAE residues $E_{10}$ (the side chain of which is disordered in this structure), $D_{16}$, $N_{38}$ and $D_{104}$. In addition, $R_{753}$ of RD could interact with $E_{197}$ or the C-terminus, highly conserved PflAE residues $N_{38}$ and $H_{202}$ are positioned to interact with RD glycine loop and RD $Y_{735}$ stacks against PflAE $L_{204}$, which is always hydrophobic (Figure III.12f). Of the residues involved in these possible interactions, $E_{10}$, $D_{16}$, $N_{38}$ and $H_{202}$ are notable in that they are fully conserved among the PFL activases, and $E_{197}$ is always either glutamate or aspartate. RD also contacts conserved surface regions 1 and 2 (Figure III.12e), providing a reasonable explanation for their preservation across the activases.

## III.D. Implications of the PflAE structures

The AdoMet radical superfamily is capable of catalyzing some of the most difficult chemical reactions known in biology via the breaking of unactivated C-H bonds to form substrate radicals[2,3]. The structures described here illustrate the surprising plasticity of the core fold and extend our understanding of the activation reactions of the GRE family. The AdoMet radical structures[5-8] show that, in general, the size of the specific enzyme tends to increase as its substrate gets smaller, a point emphasized by PflAE's small size and conspicuous lack of non-core secondary structural elements. As the most compact of the structurally characterized AdoMet radical enzymes, with 245 residues compared to next-smallest MoaA's 340 residues, PflAE defines the minimal machinery required for generation of 5'-dA•. The majority of interactions between PflAE

and its substrate occur outside of the 180-residue AdoMet radical core, as less than one half of the contacts between the PflAE and the peptide substrate originate from within the core.

The mechanism of activation of Pfl and other GREs is an open question, though six independent structures[18–23] of inactive GREs have revealed their 10-stranded $\beta$ barrel folds. Although the radical-harboring $G_{734}$ is buried 8Å from the surface of the protein[18], hydrogen atom abstraction by 5'-dA• to form the glycyl radical occurs directly and stereospecifically[24]. In order to allow access of the activase to the catalytic glycine, a dramatic conformational change within Pfl would have to occur. This ATP-independent process could proceed in two ways, the first through the action of the activase via conformational changes induced by its binding. Alternatively, Pfl itself could be conformationally flexible, at times adopting a more open conformation that exposes $G_{734}$, but which is less thermodynamically stable than the conformation observed in the crystal structures.

The findings presented here provide important insight into the possible modes of interaction in the full physiological complex and suggest that Pfl is a modular protein capable of conformational flexibility, although they do not answer the question of how the necessary conformational changes occur. Because the peptide used in these studies is known to stimulate AdoMet cleavage by PflAE[24], the pept-AE structure is relevant to the physiological reaction. The location of peptide binding and the AdoMet C5' to $G_{734}$ C$\alpha$ distance are consistent with the substrate binding sites observed in the other AdoMet radical enzyme structures[5, 8]. Additionally, the torsional angles of $G_{734}$ observed in this

structure are similar to both those seen in the Pfl structure[18, 21] and those predicted for the

Pfl glycyl radical by quantum-chemical studies of the EPR parameters of the Pfl radical[34].

The pept-AE structure illustrates one possible mode of interaction between Pfl and the

activase in which the glycine loop extends across the surface of the PflAE active site.

The unexpected motion undergone by loop A to bind substrate is likely a feature

conserved among all of the activases, as evidenced by high conservation of the DGxGxR

motif and by the conservation of the residues that come into contact with the motif after it

binds substrate. This loop could play a major role in the conformational change of Pfl by

shifting the glycine loop into the PflAE active site, and it is most certainly essential to the

proper orientation of $G_{734}$.


Through sequence comparisons and docking studies, we were able to identify a

fragment of Pfl, residues 712 – 759, that we believe is the main unit that interacts with

PflAE in the full complex. This 48-residue fragment forms a small domain that connects

to the $\beta$ barrel of Pfl through a short loop. Rotation of this "radical domain" (RD) about

a single hinge point moves the glycine loop out of the barrel, making it accessible to the

activase (Figure III.16). This motion could occur with few rearrangements of Pfl and

minimal steric clashes. Our current view of the Pfl–PflAE complex formation is that

PflAE binding either helps position RD outside of the Pfl barrel or captures a

conformationally flexible RD in its "out" conformation. With RD rotated out of Pfl to a

sufficient degree, PflAE loop A can swing up to bind and orient $G_{734}$ properly. The

resulting docking model is attractive in terms of the positioning of $G_{734}$ with respect to

the AdoMet, explaining the two regions of PflAE surface conservation, and providing a

role for conservation of residues $D_{15}$, $H_{37}$ and $N_{38}$ in RD binding. After radical

generation, RD rotates back into the Pfl barrel, likely with the help of PflAE to shield the radical. Presumably after glycyl radical formation, a conformational change of the glycine loop must serve to lock the loop in place, preventing destruction of the protein radical. After generating a radical in one monomer of the Pfl dimer, PflAE's task is complete. Unknown allosteric effects are thought to prevent activation of the second monomer, explaining the observed half-of-the-sites reactivity[18]. Since RD used in these docking studies has an identical sequence to the C-terminal 48 residues of YfiD, monomeric YfiD is expected to interact with PflAE in a similar way.

The Pfl system is responsible for a fundamental, metabolically essential chemical transformation and is believed to be representative of one of Nature's most ancient enzymes. Our PflAE structures identify the regions and residues involved in activation and complex formation with Pfl, and provide a much needed structural framework for future experiments on this system.

## III.E. Materials and Methods

### III.E.1. Crystallization of AE

PflAE was purified according to published procedures[35, 36] except that 1 mM DTT was included in all buffers. The peptide (RVSGYAV, purity $\geq$ 98 %) was obtained from Celteck Peptides. AdoMet was synthesized using AdoMet synthetase as previously described[36]. Crystals of AE were obtained in a Coy anaerobic chamber under an atmosphere of 5% $H_2$ / 95% Argon gas mix at room temperature via hanging drop vapor diffusion by mixing 1 μL protein (0.730 mM in 50 mM HEPES, 200 mM NaCl, 1 mM DTT, 7.25 mM AdoMet, pH 7.5) and 1 μL crystallization buffer (0.1 M Tris pH 8.5, 25%

PEG 3350). Drops were equilibrated against a well solution of 2.5 M ammonium sulfate to improve reproducibility and the crystallization time (typically four days). To improve diffraction quality, a detergent (0.11 mM octaethylene glycol monododecyl ether, $C_{12}E_8$, Hampton Research) was added to the crystallization drop and to all cryoprotectant solutions. Crystals were cryoprotected with a final concentration of 20% PEG 400 or 20% 2-methylpentane-2,4-diol (MPD) added to the crystallization solutions, followed by freezing in liquid nitrogen in the anaerobic chamber.

*III.E.2. Data collection and structure determination of AE*

Multiple Anomalous Dispersion (MAD) and native datasets were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamlines 9-1, 9-2 and 1-5, and at the Advanced Light Source (ALS) on beamline 5.0.2. MAD datasets were collected in 20°–30° wedges using the inverse beam technique at the iron absorption edge, inflection and a remote wavelength. The highest quality datasets are presented here (Table III.1). The data processed in mosflm[43] and HKL2000[44] scaled reasonably well in space groups C2 (three molecules per asymmetric unit, ASU), $P3_1$ (two molecules / ASU) and $P3_121$ (one molecule / ASU). The data were of poor quality as judged by $\chi^2$ and mosaicity values (which ranged between 1.0 and 1.8) and were initially difficult to index due to poor spot profiles and multiple lattices. Attempts to phase the data in $P3_121$ were unsuccessful, and the first reasonable maps were generated in $P3_1$. In the $P3_1$ space group, two iron sites (one for each cluster per molecule) were found by SOLVE[37] (www.solve.lanl.gov) using SAD data obtained with the iron peak wavelength dataset. These sites matched anomalous difference Patterson maps calculated by XtalView[39] and

90

yielded interpretable maps to 3.5 Å resolution after refinement of the iron sites in

SHARP[38] (FOM (acent/cent) = 0.65 / 0.48) and solvent flattening in SOLOMON[45].


*III.E.3. Initial model building and refinement of AE using MAD data*

A model comprising two molecules (residues 5 to 230) was built in XtalView into

solvent-flattened experimental maps calculated to 3.5 Å resolution. After building the

cluster binding loop and placing the cluster ligands into density, the individual iron sites

were fit and refined to calculate maps to 2.87 Å resolution. The higher resolution maps

were used to further complete the model, and approximately 40% of the residue side

chains were added. Two chain breaks per molecule were observed, the first at residues

46-48 and a second break of unknown length toward the C-terminal end of each chain.

The side chains of residues at the C-terminus could not be modeled unambiguously.

High R factors and overfitting of the model in CNS[40], possible signs of twinning, led to

use of the 1 0 0 -1 -1 0 0 0 -1 twin law (which corresponds to the h,-h-k,-l symmetry

operator) in SHELXL[46] with a twin fraction of 0.5 (perfect merohedral twinning), though

the intensity distribution of the data was normal and did not indicate twinning in any of

the twin tests used. Refinement by this strategy yielded encouraging R factors (R =

28.82, $R_{free}$ = 34.69), though no further improvement was possible. These problems

(possible twinning and poor data quality) initially prevented the complete refinement of

the model. A second crystal form was then identified, allowing us to surmount this

problem.


*III.E.4. Crystallization of pept-AE*

Substrate-bound PflAE was also crystallized at room temperature using hanging drop vapor diffusion in an anaerobic chamber. The initial conditions were identified by sitting drop vapor diffusion using high throughput trays and the Hampton Index Screen. Crystals formed in approximately six days from drops comprised of 1 μL protein (0.68 mM protein, 6.80 mM AdoMet, 6.95 mM peptide, 50 mM HEPES, 200 mM NaCl, and 1 mM DTT) and 1 μL crystallization buffer (100 mM HEPES pH 6.8, 3.5 M sodium formate). Larger crystals were obtained after optimization of the original condition by addition of detergents (Detergent Screens, Hampton Research). A single drop containing 0.2 mM 2,6-dimethyl-4-heptyl-β-D-maltopyranoside as an additive yielded several large (80 μm × 80 μm × 100 μm) crystals of good enough quality for data collection. These crystals were cryocooled with liquid nitrogen directly from the drop in the anaerobic chamber and stored for synchrotron data collection.

### III.E.5. Data collection and structure determination of pept-AE

Native and MAD datasets were collected on beamlines 9-1 and 9-2 at SSRL. The data were processed with HKL2000 in space group $P6_122$, with 1 molecule per ASU and a solvent content of 54.2% (Table III.1). Because molecular replacement with the model built from the first crystal form did not yield a solution that refined well, MAD data was used to obtain experimental maps. One site corresponding to the iron sulfur cluster was identified with SOLVE and refined in SHARP. The resulting maps were solvent flattened in SOLOMON (through SHARP) to give interpretable density to 2.77 Å resolution (FOM (acent/cent) = 0.47 / 0.33).

### III.E.6. Model building and refinement of pept-AE

The model built from the first crystal form was fit into the new experimental maps manually, followed by rigid-body refinement in CNS. The new maps allowed placement of most side chains and correction and placement of the entire polypeptide backbone (residues 2-245), with no chain breaks, as well as placement of the cluster, AdoMet and peptide substrate. Calculation of phase combined $2F_o$-$F_c$ and $F_o$-$F_c$ maps in CNS using the MLHL target allowed confident assignment of the peptide substrate orientation. Refinement was carried out in CNS with no sigma cutoff to $R_{work}$ = 22.9 and $R_{free}$ = 26.1 with 5.2% of reflections in the test set (Table III.2). The final model contains PflAE residues 2-256, 4 iron atoms, 4 sulfur atoms, 27 AdoMet atoms, six peptide substrate residues, 16 waters and five formate molecules (10 atoms) from the crystallization conditions.

### III.E.7. Final model building and refinement of AE using pept-AE structure

After the pept-AE model was refined to reasonable R factors, these coordinates, minus the peptide substrate, were used to refine against data from the first crystal form. Refinement was carried out in CNS and SHELXL in C2, $P3_1$ and $P3_121$ to determine the proper space group. Fortunately, refinement using the new (corrected) model in CNS in $P3_121$ (with no twin law) gave better R factors and improved electron density maps. Apparently the initial experimental maps were simply too poor to correctly build parts of the model, in particular the loop following β6, which extends 25 Å over the cluster from the C-terminal end of the β sheet and includes 25 residues. Rounds of manual rebuilding and refinement in CNS with no sigma cutoff were done to a final $R_{work}$ and $R_{free}$ of 23.9 and 32.6, respectively, using a test set of 9.5% of reflections (Supplementary Table 2).

This final model contains residues 1-245 of PflAE, 4 iron atoms, 4 sulfur atoms, and 49 waters.

We have confirmed in several ways that our AE data is indeed not twinned. First, there was no indication of twinning in the cumulative intensity distribution of the AE data (the following programs and servers were used to confirm: Merohedral Crystal Twinning Server (http://nihserver.mbi.ucla.edu/Twinning/)[47], the Merohedral Twin Detector: Padilla-Yeates Algorithm (http://nihserver.mbi.ucla.edu/pystats/)[48], Scala[49] in ccp4[50] and CNS). Further tests on the AE data processed in $P3_121$ did not indicate any twinning in that space group. Second, it is unlikely that a strong anomalous signal would be detected – and then used to successfully generate reasonable experimental maps – in perfectly twinned data. Third, the self rotation functions calculated in for this data are completely consistent with the space group $P3_121$. Fourth, the twin operator (h,-h-k,-l, equivalent to k,h,-l) used for refinement in Shelxl is actually a symmetry operator for space group $P3_121$. Finally, we were able to obtain lower R factors with the corrected model with normal refinement in CNS. We therefore attribute the overfitting observed in our final R factors and the high $R_{free}$ to poor data quality, and limit our discussion of the AE model as a result.

*III.E.8. Model preparation for docking studies*

The AE and pept-AE models were docked with a fragment of the PFL model (RCSB code 2PFL)[18] corresponding to the portion of PFL homologous to the small protein YfiD[27] (residues 700 – 759). This fragment, termed the PFL radical domain (RD), was further truncated at the N-terminus to improve docking results and minimize

steric clashes. When the pept-AE was used in docking studies, the peptide residues were deleted from the model before manipulation.

*III.E.9. Manual docking*

A manual docking model of the RD with pept-AE was generated by orientation of the glycine loop of the RD in the active site in as close a conformation as possible to that of the bound peptide (see Figure III.1). This docking model (Figure III.6) was used as a guide to help evaluate the complexes output by ZDOCK[41].

*III.E.10. Docking with ZDOCK*

The docking algorithm ZDOCK generates the possible binding modes of two proteins based on their shape complementarity, electrostatics and desolvation energies[41]. ZDOCK was used as described on the program's website (http://zlab.bu.edu/zdock/). Residues of pept-AE and the RD were blocked from evaluation by the program if their involvement in the catalytically relevant complex was not expected (i.e., residues located on the opposite side of the AE with respect to the active site, see Figure III.14).

*III.E.11. Evaluation of docking models*

The top 20 complexes as scored by ZDOCK from each of six separate trials were evaluated based on the position of Pfl $G_{734}$ with respect to the AE active site and distance between $G_{734}$ C$\alpha$ and AdoMet C5'. Also, the degree of conservation of residues on the surface of PflAE, calculated using ESPript[42], was used to evaluate the docking models. The eight most reasonable models were then refined using rigid-body refinement and minimization in CNS (with no X-ray term) and reevaluated based on the criteria

mentioned above as well as possible hydrogen bonding contacts, interaction of RD with conserved AE residues, and agreement with biochemical data on the involvement of specific Pfl residues in complex formation. In particular, $R_{731}$ of Pfl is necessary for the Pfl activation reaction, while the hydroxyl groups of $S_{733}$ and $Y_{735}$ are not required[24]. The model reported here (ZDOCK score of 31.12) was selected as the best complex based on the criteria discussed above, although it was not the highest-scored in the automated docking algorthims.

*III.E.12. Equipment and Settings*

All figures were generated with Pymol[51] with the exception of Figure III.8, which was generated with ClustalW[52] and Weblogo[53]. Adobe Photoshop was used to add labels to all figures.

## IV.F. References

1    Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **29**, 1097-1106 (2001).

2    Wang, S. C. & Frey, P. A. S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem. Sci.* **32**, 101-110 (2007).

3    Cheek, J. & Broderick, J. B. Adenosylmethionine-dependent iron-sulfur enzymes: versatile clusters in a radical new role. *J. Biol. Inorg. Chem.* **6**, 209-226 (2001).

4    Blaschkowski, H. P., Neuer, G., Ludwig-Festl, M. & Knappe, J. Routes of flavodoxin and ferredoxin reduction in Escherichia coli. CoA-acylating pyruvate: flavodoxin and NADPH: flavodoxin oxidoreductases participating in the activation of pyruvate formate-lyase. *Eur. J. Biochem.* **123**, 563-569 (1982).

5    Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* **303**, 76-69 (2004).

[6]     Layer, G., Moser, J., Heinz, D. W., Jahn, D. & Schubert, W. D. Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J.* **22**, 6214-6224 (2003).

[7]     Hanzelmann, P. & Schindelin, H. Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12870-12875 (2004).

[8]     Lepore, B. W., Ruzicka, F. J., Frey, P. A. & Ringe, D. The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13819-13824 (2005).

[9]     Rebeil, R. et al. Spore photoproduct lyase from Bacillus subtilis spores is a novel iron-sulfur DNA repair enzyme which shares features with proteins such as class III anaerobic ribonucleotide reductases and pyruvate-formate lyases. *J. Bacteriol.* **180**, 4879-4885 (1998).

[10]    Tamarit, J., Mulliez, E., Meier, C., Trautwein, A. & Fontecave, M. The anaerobic ribonucleotide reductase from Escherichia coli. The small protein is an activating enzyme containing a [4fe-4s](2+) center. *J. Biol. Chem.* **274**, 31291-31296 (1999).

[11]    Selmer, T., Pierik, A. J. & Heider, J. New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol. Chem.* **386**, 981-988 (2005).

[12]    Raynaud, C., Sarcabal, P., Meynial-Salles, I., Croux, C. & Soucaille, P. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5010-5015 (2003).

[13]    Leuthner, B. et al. Biochemical and genetic characterization of benzylsuccinate synthase from Thauera aromatica: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol. Microbiol.* **28**, 615-628 (1998).

[14]    Yu, L., Blaser, M., Andrei, P. I., Pierik, A. J. & Selmer, T. 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* **45**, 9584-9592 (2006).

[15]    Reichard, P. The evolution of ribonucleotide reduction. *Trends Biochem. Sci* **22**, 81-85 (1997).

[16]    Stubbe, J. Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr. Opin. Struct. Biol.* **10**, 731-736 (2000).

[17]    Knappe, J. & Wagner, A. F. V. Glycyl Free Radical in Pyruvate Formate-Lyase: Synthesis, Structure Characteristics, and Involvement in Catalysis. *Methods Enzymol.* **258**, 343-361 (1995).

[18]    Becker, A. et al. Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nature Struct. Biol* **6**, 969-975 (1999).

[19]    Logan, D. T., Andersson, J., Sjöberg, B. M. & Nordlund, P. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283**, 1499-1504 (1999).

[20] Leppänen, V. M. et al. Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure* **7**, 733-744 (1999).

[21] Lehtiö, L., Leppänen, V. M., Kozarich, J. W. & Goldman, A. Structure of Escherichia coli pyruvate formate-lyase with pyruvate. *Acta Crystallogr.* **D59**, 2209-2212 (2002).

[22] O'Brien, J. R. et al. Insight into the mechanism of the B12-independent glycerol dehydratase from Clostridium butyricum: preliminary biochemical and structural characterization. *Biochemistry* **43**, 4635-4645 (2004).

[23] Lehtiö, L., Grossmann, J. G., Kokona, B., Fairman, R. & Goldman, A. Crystal structure of a glycyl radical enzyme from Archaeoglobus fulgidus. *J. Mol. Biol.* **357**, 221-235 (2006).

[24] Frey, M., Rothe, M., Wagner, A. F. V. & Knappe, J. Adenosylmethionine-dependent Synthesis of the Glycyl Radical in Pyruvate Formate-lyase by Abstraction of the Glycine C-2 pro-S Hydrogen Atom. *J. Biol. Chem.* **269(17) pp 12432-12437** (1994).

[25] Knappe, J., Neugebauer, F. A., Blaschkowski, H. P. & Ganzler, M. Post-Translational Activation Introduces a Free Radical into Pyruvate Formate-Lyase. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1332-1335 (1984).

[26] Wagner, A. F. V., Frey, M., Neugebauer, F. A., Schafer, W. & Knappe, J. The Free Radical in Pyruvate Formate-Lyase is Located on Glycine-734. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 996-1000 (1992).

[27] Wagner, A. F. V. et al. YfiD of Escherichia coli and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem. Biophys. Res. Commun.* **285**, 456-462 (2001).

[28] Walsby, C. J., Ortillo, D., Broderick, W. E., Broderick, J. B. & Hoffman, B. M. An anchoring role for FeS clusters: chelation of the amino acid moiety of S-adenosylmethionine to the unique iron site of the [4Fe-4S] cluster of pyruvate formate-lyase activating enzyme. *J. Amer. Chem. Soc.* **124**, 11270-11271 (2002).

[29] Walsby, C. J. et al. Electron-nuclear double resonance spectroscopic evidence that S-adenosylmethionine binds in contact with the catalytically active $[4Fe-4S]^{(+)}$ cluster of pyruvate formate-lyase activating enzyme. *J. Amer. Chem. Soc.* **124**, 3143-3151 (2002).

[30] Walsby, C. J. et al. Spectroscopic approaches to elucidating novel iron-sulfur chemistry in the "radical-Sam" protein superfamily. *Inorg. Chem* **44**, 727-741 (2005).

[31] Nicolet, Y. & Drennan, C. L. AdoMet radical proteins--from structure to evolution--alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res.* **32**, 4015-4025 (2004).

[32] Krebs, C., Broderick, W. E., Henshaw, T. F., Broderick, J. B. & Huynh, B. H. Coordination of Adenosylmethionine to a Unique Iron Site of the [4Fe-4S] of

Pyruvate Formate-Lyase Activating Enzyme: A Mössbauer Spectroscopic Study. *J. Amer. Chem. Soc.* **124**, 912-913 (2002).

33   Wong, K. K. et al. Molecular properties of pyruvate formate-lyase activating enzyme. *Biochemistry* **32**, 14102-14110 (1993).

34   Kacprzak, S., Reviakine, R. & Kaupp, M. Understanding the electon paramagnetic resonance parameters of protein-bound glycyl radicals. *J. Phys. Chem. B.* **111**, 820-831 (2007).

35   Broderick, J. B. et al. Pyruvate formate-lyase-activating enzyme: strictly anaerobic isolation yields active enzyme containing a [3Fe-4S](+) cluster. *Biochem. Biophys. Res. Commun.* **269**, 451-456 (2000).

36   Walsby, C. J. et al. Electron-Nuclear Double Resonance Spectroscopic Evidence That S-Adenosylmethionine Binds in Contact with the Catalytically Active [4Fe-4S]+ Cluster of Pyruvate Formate-Lyase Activating Enzyme. *J. Amer. Chem. Soc.* **124**, 3143-3151 (2002).

37   Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr.* **D55**, 849-861 (1999).

38   Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr.* **D59**, 2023-2030 (2003).

39   McRee, D. E. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Bio.* **125**, 156-165 (1999).

40   Brunger, A. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905-921 (1998).

41   Chen, R., Li, L. & Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **52**, 80-87 (2003).

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

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

44   Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 307-326 (1997).

45   Abrahams, J. P. & Leslie, A. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr.* **D52**, 30-42 (1996).

46   Sheldrick, G. & Schneider, T. SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319-343 (1997).

47   Yeates, T. O., Detecting and overcoming crystal twinning. *Methods Enzymol* **276**, 344 (1997).

[48] Padilla, J. E. and Yeates, T. O., A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr D Biol Crystallogr* **59** (Pt 7), 1124 (2003).

[49] Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62** (Pt 1), 72 (2006).

[50] The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50** (Pt 5), 760 (1994).

[51] DeLano, W. The PyMOL Molecular Graphics System. {http://www.pymol.org} (2002).

[52] Claverie, J. & States, D. Information enhancement. methods for large scale sequence analysis. *Comput. Chem.* **17**, 191-201 (1993).

[53] Crooks, G., Hon, G., Chandonia, J. & Brenner, S. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

### III.G. Tables and Figures

### Table III.1: Data Collection Statistics.

| Dataset | AE | | | | pept-AE | | | |
|---|---|---|---|---|---|---|---|---|
| | Peak | Inflection | Remote | Native | Peak | Inflection | Remote | Native |
| Wavelength (Å) | 1.73542 | 1.74166 | 1.37755 | 1.00000 | 1.73955 | 1.74158 | 1.36241 | 1.00000 |
| Space Group | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Resolution (Å) | 2.87 | 3.20 | 3.70 | 2.25 | 2.90 | 2.80 | 2.70 | 2.80 |
| Rsymm (%) | 7.1(30.5) | 7.0(27.8) | 7.9(26.0) | 7.8(32.2) | 9.8(42.0) | 8.1(36.9) | 7.3(35.9) | 5.2(35.6) |
| Unique Obs | 9997(1021) | 7363(742) | 4680(460) | 11183(1103) | 12853(1152) | 12677(1107) | 14585(1406) | 8085(731) |
| I/sigma | 36.9(7.4) | 37.6(8.3) | 21.1(7.1) | 23.9(4.0) | 18.6(2.7) | 13.1(3.1) | 19.9(3.1) | 25.4(3.63) |
| %Complete | 100.0(100.0) | 99.9(100.0) | 99.8(100.0) | 99.0(93.7) | 98.5(87.6) | 88.7(78.6) | 98.6(95.8) | 98.7(94.3) |
| Redundancy | 10.8(10.9) | 10.7(10.8) | 5.0(4.9) | 9.6(4.4) | 9.8(5.8) | 5.4(5.0) | 5.5(4.0) | 5.5(4.4) |
| a | 58.09 | 58.25 | 58.27 | 57.98 | 74.56 | 74.41 | 74.43 | 74.35 |
| c | 117.27 | 117.64 | 117.85 | 117.37 | 187.98 | 187.46 | 187.61 | 187.45 |

**Values in parenthesis refer to the high-resolution bin.**

**Table III.2: Refinement Statistics.**

| | AE | Pept-AE |
|---|---|---|
| Resolution Limits (Å) | 38.19 - 2.25 | 29.13 - 2.77 |
| # Unique Reflections | 19688 | 8376 |
| # Reflections in Test Set | 1869 | 438 |
| $R_{work}$ (%)[a] | 23.9 | 22.9 |
| $R_{free}$ (%) | 32.3 | 26.1 |
| Average B factor ($\text{Å}^2$) | 56.3 | 80 |
| Protein | 55.7(1902) | 76.9(1872) |
| 4Fe-4S | 44.3(8) | 59.7(8) |
| AdoMet | N/A | 75.7(27) |
| Peptide | N/A | 104.4(42) |
| Water | 59.0(49) | 81.0(16) |
| Bond length deviation (Å) | 0.006 | 0.008 |
| Bond angle deviation (°) | 1.4 | 1.5 |

**Values in parenthesis refer to number of atoms.**

**Table III.3: RMSDs between structures of AdoMet radical proteins.**

|      | Pept-AE    | AE         |
|------|------------|------------|
| AE   | 0.77 (245) | ———        |
| BioB | 2.04 (100) | 2.08 (98)  |
| HemN | 1.99 (109) | 2.01 (115) |
| MoaA | 1.56 (145) | 1.47 (140) |
| LAM  | 2.05 (101) | 1.97 (107) |

RMSDs are shown in Å. The number of Cα atoms matched in the calculation is shown in parenthesis.

**Figure III.1: Stereoview of AE active site** with the 4Fe-4S cluster and protein sidechains shown in sticks and colored as follows: iron, ruby; sulfur, gold; carbons, grey; oxygens, red and nitrogens, blue. Water molecules are shown as red spheres. A $2F_o$-$F_c$ electron density map calculated by CNS and contoured at $1\sigma$ is shown in blue mesh.
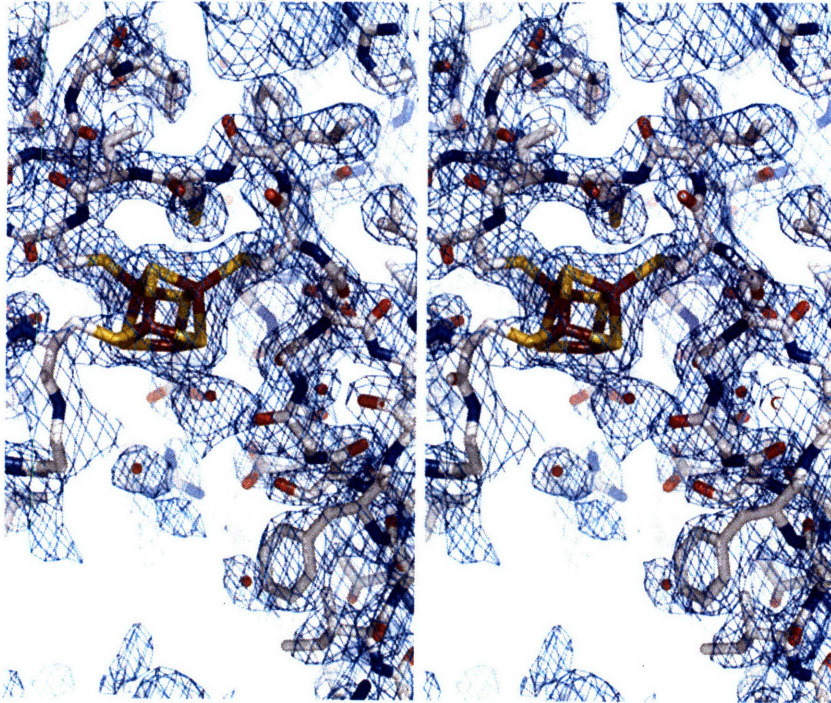
**Figure III.2: Stereoview of pept-AE active site** with the 4Fe-4S cluster, AdoMet and peptide and protein sidechains shown in sticks and colored as in Figure III.1 with the following exceptions: AdoMet carbons are colored green, and peptide substrate carbons are teal. A phase-combined $2F_o$-$F_c$ electron density map calculated by CNS and contoured at $1\sigma$ is shown in blue mesh.
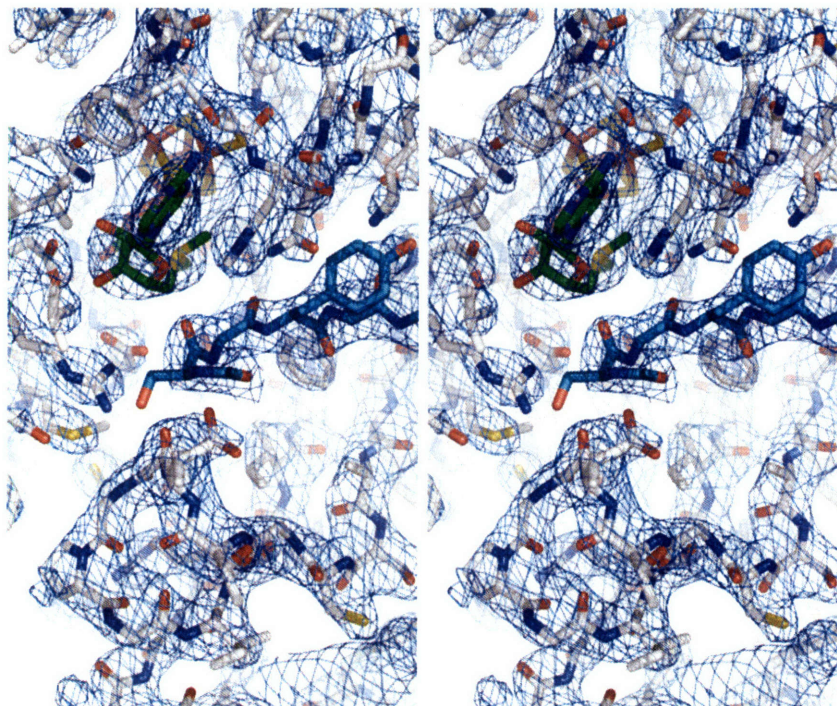
**Figure III.3: Stereoview of PflAE monomer** with secondary structural elements assigned numerically (helices in cyan, strands in yellow). The loops following strands β1′, β1, and β6 are labeled A (red, residues 10-20), B (purple, residues 27 - 47) and C (orange, residues 201-225). The 4Fe-4S cluster (ruby and gold), AdoMet (green carbons) and peptide (teal carbons) are depicted in sticks with oxygens colored red and nitrogens colored blue.
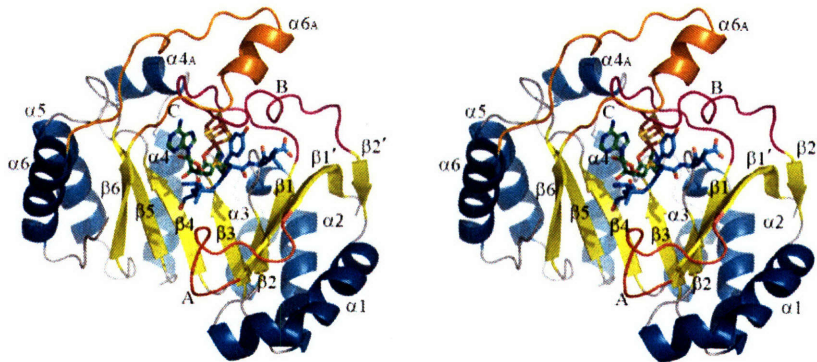
**Figure III.4: Substrate and cofactor binding.** Colors are as in Figure III.3, with protein side chain carbons in grey. Conserved motifs[32] are labeled in blue. Composite omit maps are shown as a blue mesh and are contoured at 1σ. Hydrogen bond lengths and other distances are represented as red dashed lines. **a** Detail of cluster – AdoMet interaction with composite omit map contoured around the AdoMet. Distances of interest between the unique iron of the 4Fe-4S cluster and AdoMet atoms are shown. **b** AdoMet – protein interactions. $R_{166}$ and $D_{104}$ do not make hydrogen bonds to either substrate, but may be involved in catalysis. **c** Peptide – protein interactions. **d** Omit map contoured around the peptide.
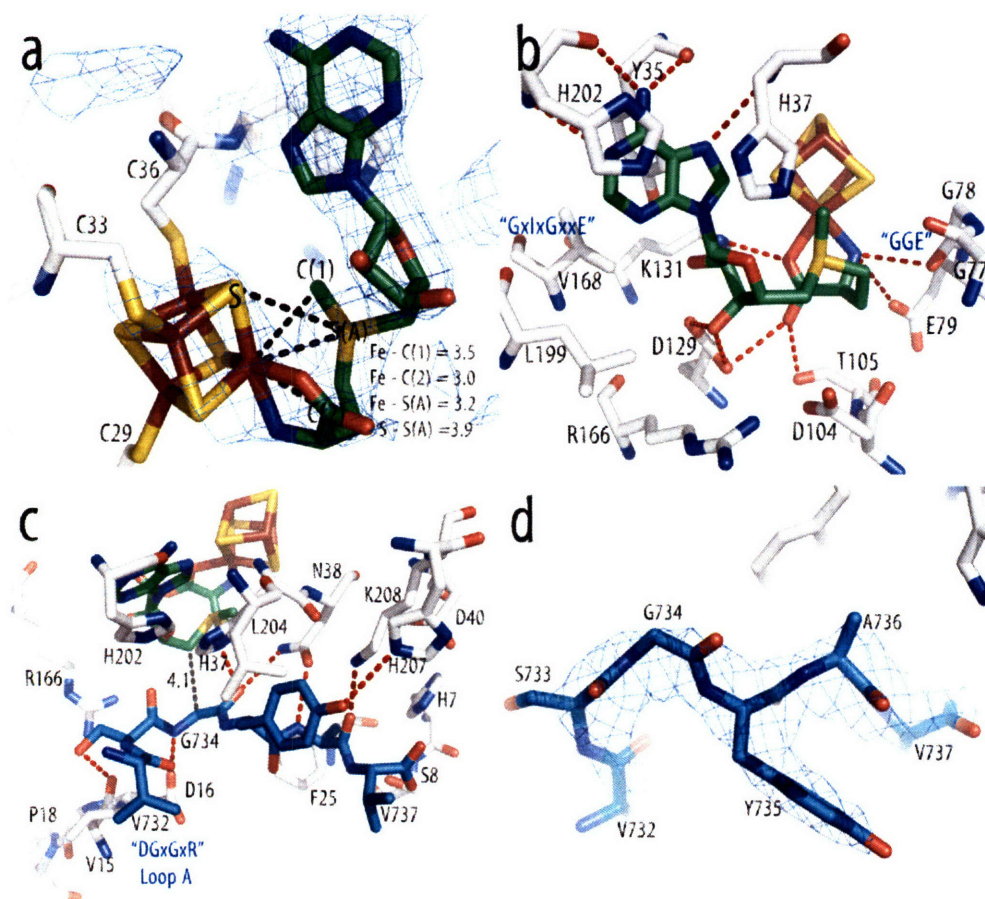
**Figure III.5: Stereoview of active site showing detail of AdoMet – peptide positioning**. The 4Fe-4S cluster, AdoMet and the peptide are shown in sticks colored as follows: iron, ruby; sulfur, gold; AdoMet carbons, green; peptide carbons, teal. The protein surface is shown colored by atom type (protein carbons, light grey; oxygens, red; nitrogens, blue).
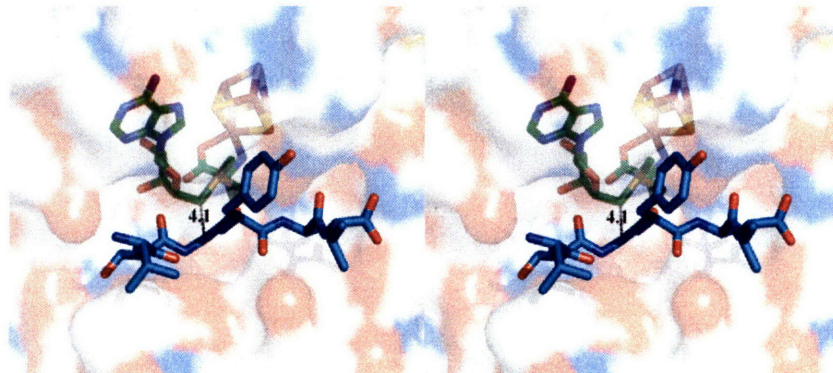
**Figure III.6: Surface representation of pept-AE** showing that no AdoMet (green) or 4Fe-4S cluster atoms (ruby and gold) are solvent exposed when substrate is bound. The bound peptide is in teal carbons with AE carbons in grey, oxygens in red and nitrogens in blue.
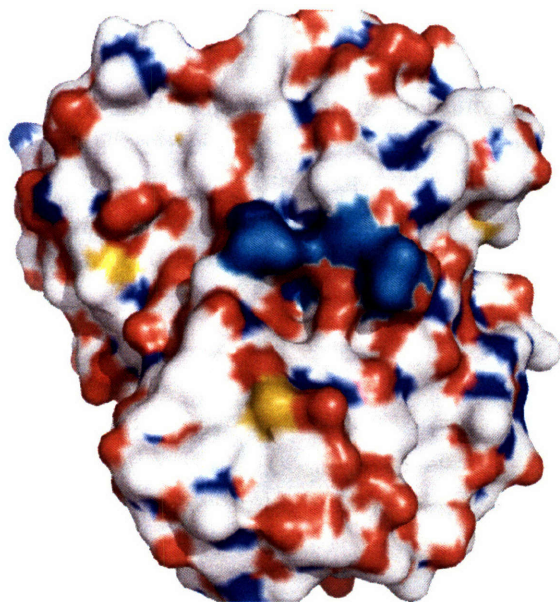
**Figure III.7: Comparison of the electron density maps of the substrate-free and substrate-bound PflAE models at the AdoMet binding site.** Colors are as in Supplementary Fig. 2. **a** The pept-AE model is shown with a phase-combined $2F_o$-$F_c$ electron density map calculated by CNS and contoured at $1\sigma$. **b** The AE model with a $2F_o$-$F_c$ electron density map calculated by CNS and contoured at $1\sigma$ is shown. To visualize the AdoMet binding site, we show AdoMet from the pept-AE model superimposed on the AE model. There is clearly no density for AdoMet in the AE model electron density maps, beyond a relatively small amount of density at the 4Fe-4S unique iron, which may result either from AdoMet interacting with the cluster without becoming well ordered in its binding site or from undefined buffer atoms. We have modelled this density as a water. It does not appear that the substrate binding site is occluded by components of the crystallization conditions, though this is a distant possibility and all crystal lattice contacts are located far from this site.
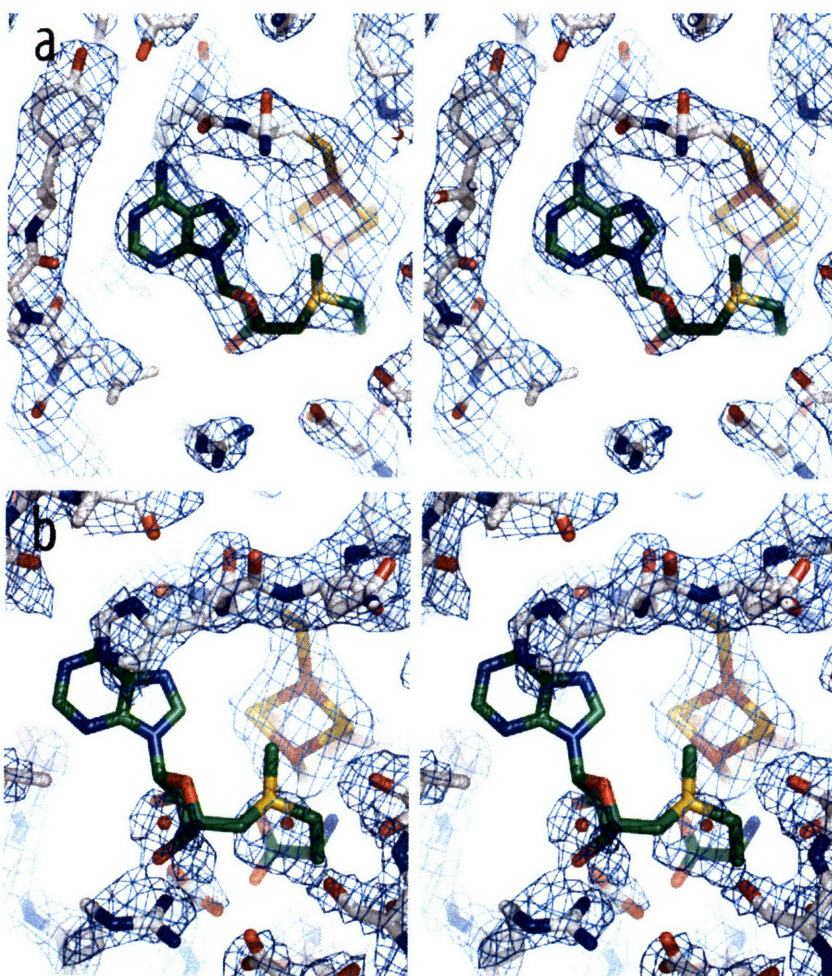
**Figure III.8: Glycyl radical enzyme consensus motifs**. Alignments were performed with ClustalW[48] using 15-16 sequences for each GRE subfamily, and logos were generated using WebLogo[49]. The letter size represents the frequency at which the residue is observed at that position in the sequence. Basic residues are colored in blue, acidic residues in red, and others in black. Consensus motifs shown are: **a** Pfl. **b** Class III RNR. Residues at position 2 are Val and Thr. **c** Glycerol dehydratase. Residues at position 7 are Ala, Asp and Val. **d** 4-Hydroxyphenylacetate decarboxylase. Residues at position 7 are Ala, Asp and Gly. **e** Benzylsuccinate synthase. **f** Overall GRE. Residues at position 2 are Val and Thr; at position 7 are Ala, Val, Gly, Asp.
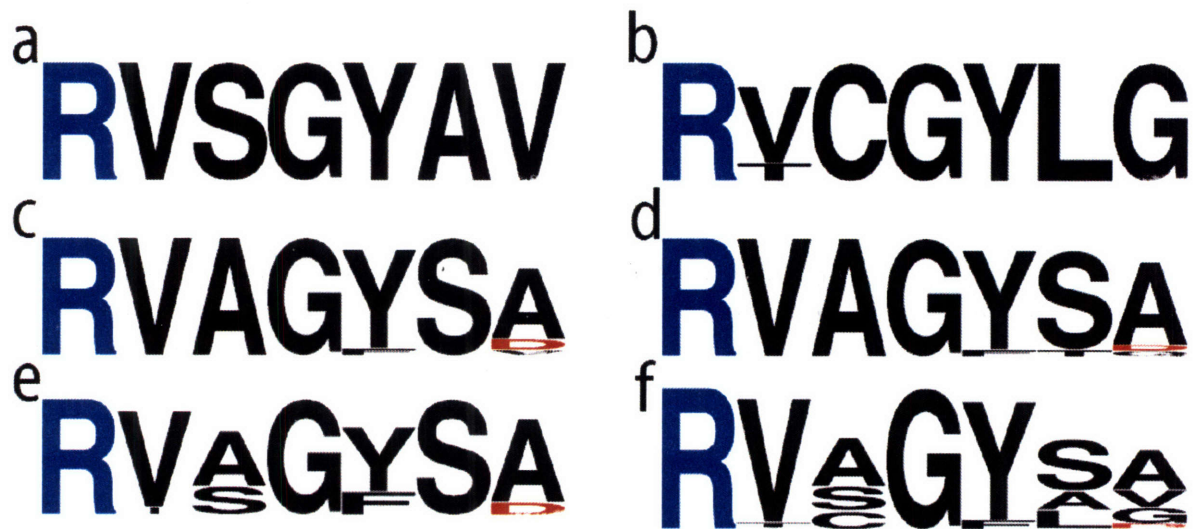
**Figure III.9: Conformational change of PflAE upon substrate binding**. Colors are as in Figure III.3 unless otherwise noted. **a** The AE and pept-AE models were superimposed and colored in gray, with areas undergoing a conformational change highlighted (blue, AE; red, pept-AE). Loops A, B and C are labeled. **b** Close-up view of the active site showing side chain rearrangements upon substrate binding. AE carbon atoms are colored blue, pept-AE carbon atoms red, oxygen red, and nitrogen blue.
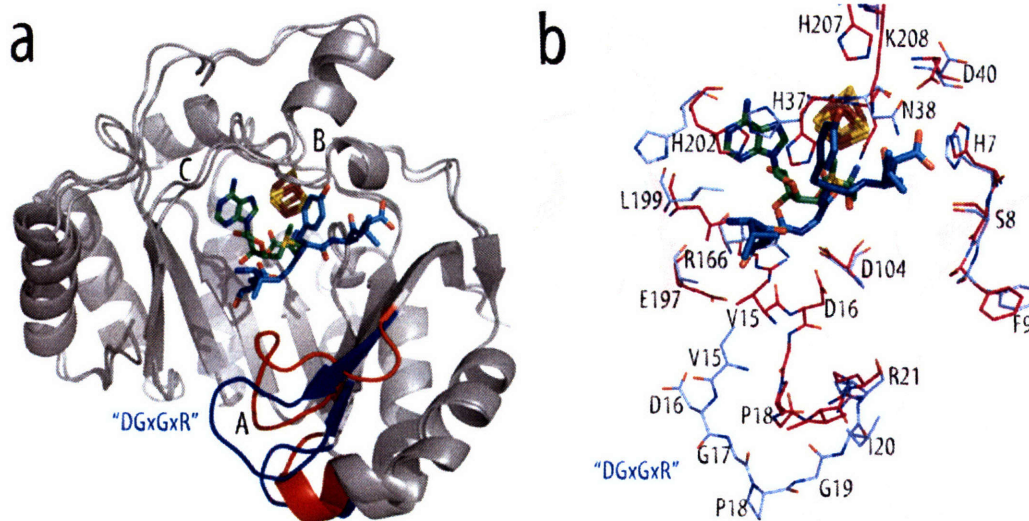
**Figure III.10: Conformation of Loop A in the AE model.** This figure shows two views of LoopA with 2Fo-Fc composite omit density calculated in CNS and contoured here at 1σ. The AE model is displayed in sticks and colored as in Figure III.1, with the exception of the Loop A residues which for clarity have yellow carbons. For comparison, Loop A from the pept-AE model is displayed as lines, with carbons colored magenta.
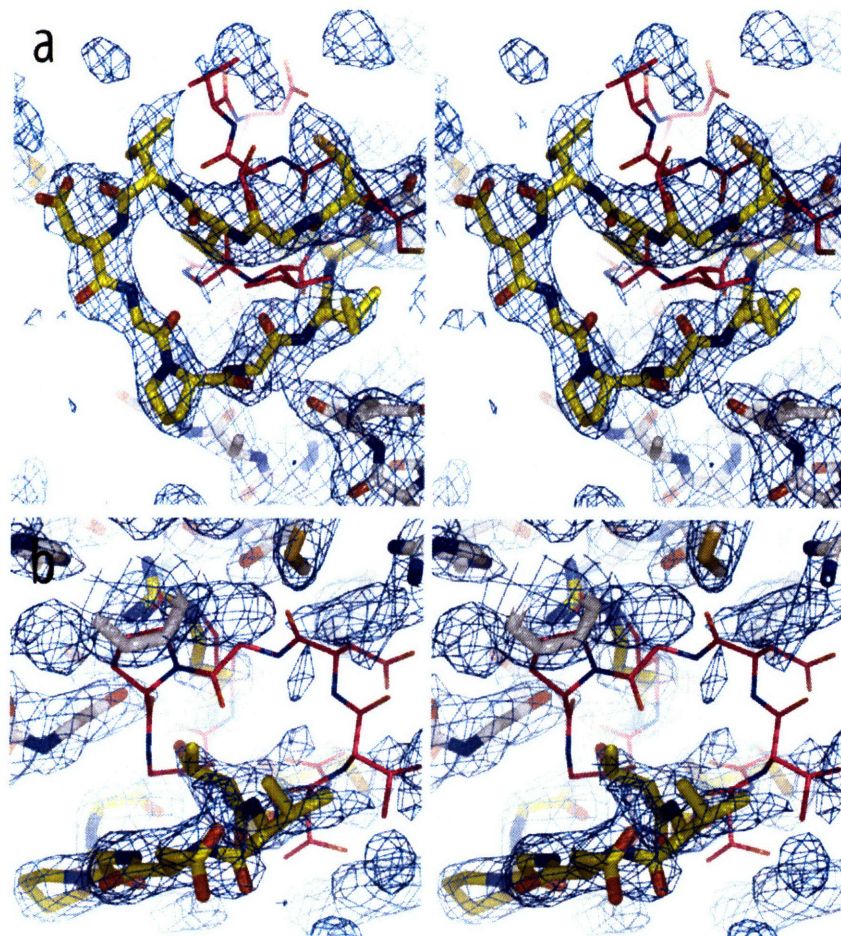
**Figure III.11: Conformation of Loop A in the pept-AE model.** This figure shows LoopA from the pept-AE model in a similar orientation as in Figure III.2, with 2Fo-Fc composite omit density calculated in CNS and contoured here at 1σ. The AE model is displayed in sticks and colored as in Figure III.2, with the exception of the Loop A residues which for clarity have magenta carbons. For comparison, Loop A from the AE model is displayed as lines, with carbons colored yellow.
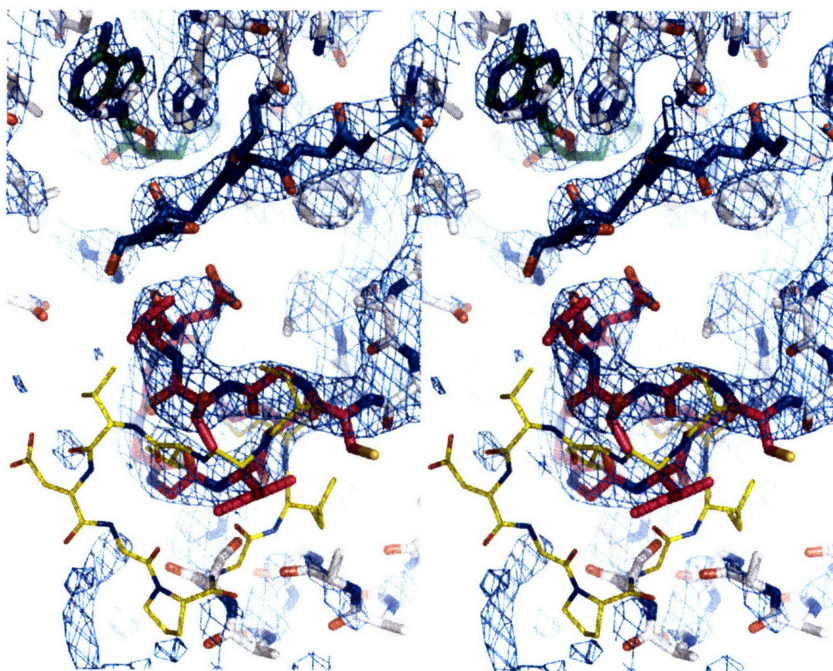
**Figure III.12: Docking studies of PflAE.** Colors are as in Figure III.3 unless specified. **a** Surface representation of the pept-AE colored by sequence conservation as calculated by ESPript[42], showing the same view of the pept-AE as in Figure III.3 and rotated 180° from that orientation. Sequence conservation is represented in rainbow colors, with areas of 100% conservation in red, and 0% conservation in blue. The three conserved regions are numbered. **b** Topology diagram of PFL showing 10-stranded α/β barrel (strands in yellow) with RD highlighted in pink and helices denoted 1 (PFL residues 712-720) and 2 (744-752). An N-terminal domain is omitted for clarity. **c** Cartoon representation of the PFL dimer in grey with RD in pink. **d** Best docking model output by ZDOCK, with Cα of G₇₃₄ displayed in spacefill and RD colored as in b. **e** Docking model, with RD displayed as in d. The pept-AE is colored as in a, and is shown in the same orientation. An arrow indicates a loop from RD. **f** Detail of active site of docking model colored as in d, with side chains of interest shown in sticks. RD side chains are labeled in magenta, pept-AE side chains in black. Dashed lines indicate possible side chain interactions between pept-AE and RD.
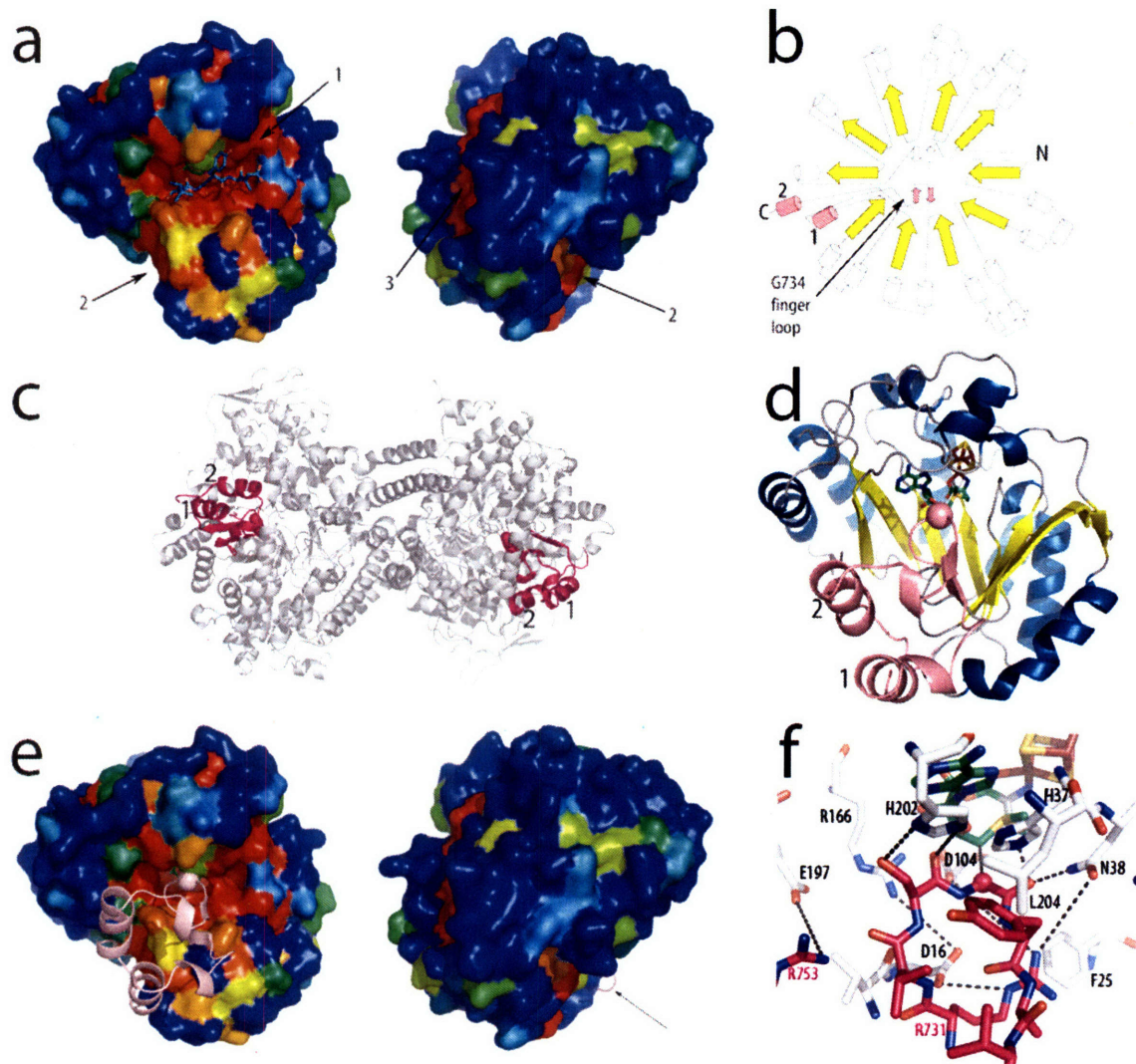
**Figure III.13: Manual docking model.** Pept-AE is colored as in Figure III.3, with the RD displayed as magenta cartoons. The peptide is shown in sticks for comparison.
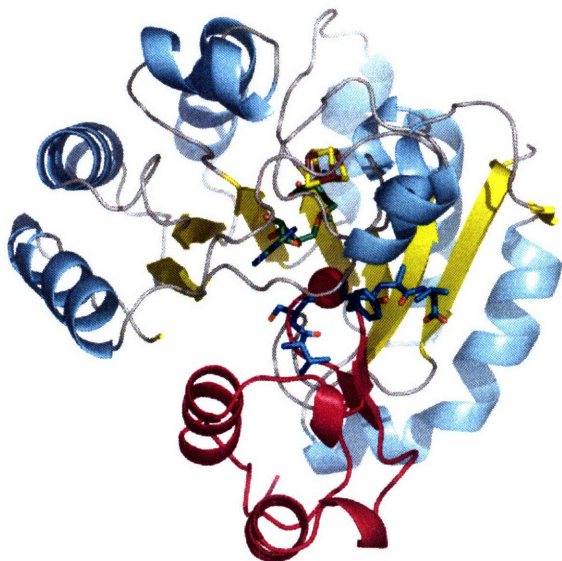
**Figure III.14: Surface representation of pept-AE colored as in Figure III.1, with residues blocked from the docking calculations highlighted in green.** Orientations are as in Figure III.9.
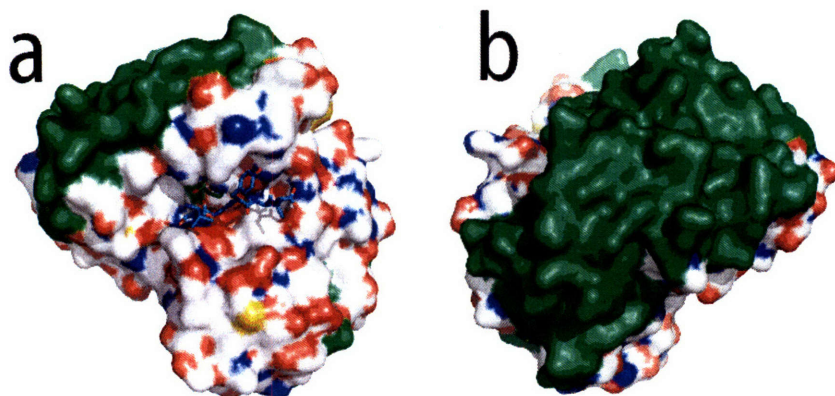


**Figure III.15: Comparison of the conformation of residues 732-737** in the pept-AE model (teal carbons) and the docked RD domain (magenta carbons). To make this figure, the best docking model was superimposed with the pept-AE model. The $C\alpha$ of each $G_{734}$ is shown as a small sphere.
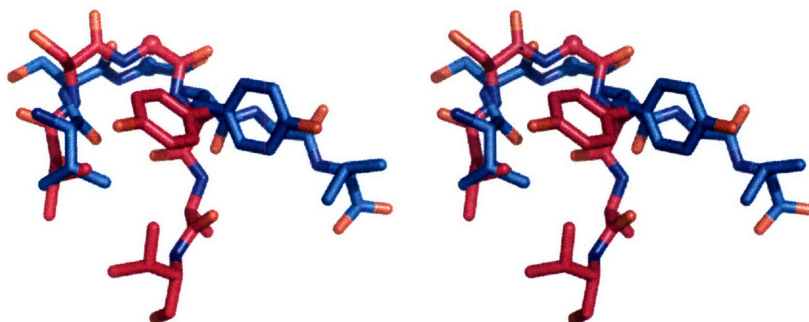
**Figure III.16: Proposed Pfl conformational change**. Stereoview of the Pfl dimer (PDB code 2PFL) showing close-up view of one monomer and focusing upon the RD domain in two orientations. The orientation in **a** is rotated approximately 60° about a horizontal axis with respect to **b**. The RD is colored light pink in its native conformation and magenta in the proposed accessible conformation. The other Pfl finger loop containing the active site cysteines ($C_{418}$ and C419) is colored yellow. The C$\alpha$ atoms of $G_{734}$, $C_{418}$ and $C_{419}$ are shown in spheres.
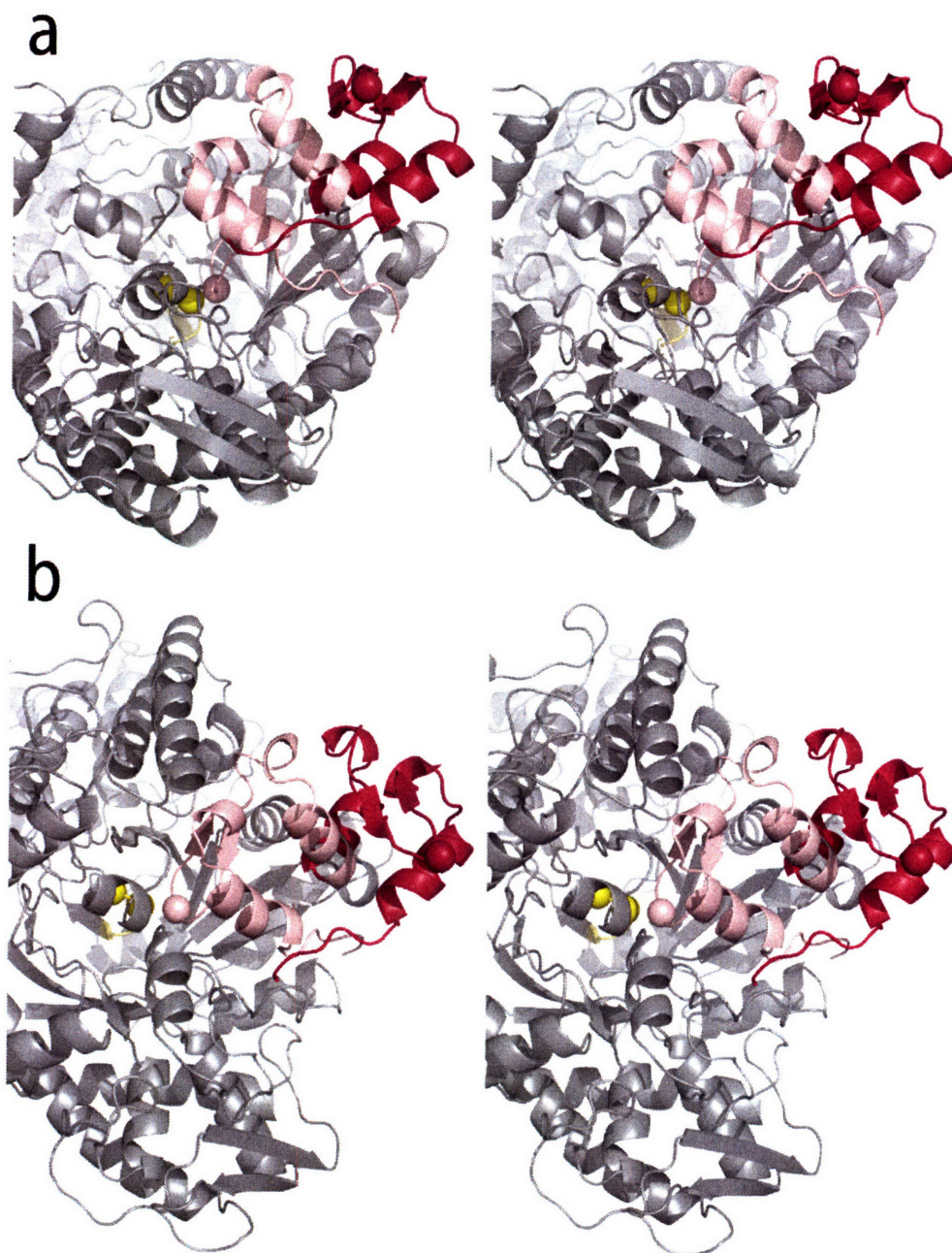
# Chapter IV:  Structural Insights into Radical Generation by the AdoMet Radical Superfamily.

## IV.A. Summary

Recent years have brought many advances in our understanding of radical generation by the AdoMet radical enzyme superfamily.  These enzymes utilize a 4Fe-4S cluster and S-adenosylmethionine to produce the 5'-deoxyadenosyl radical intermediate, which they are then able to use to catalyze substrate-specific reactions, such as sulfur insertions, the generation of a stable protein-bound glycyl radical, eliminations and rearrangements. The recent structural characterizations of five AdoMet radical enzymes have provided a wealth of information about the molecular basis for catalysis by this superfamily.  The structures now available encompass very disparate chemical reactions, which is reflected in the structural diversity of their overall three dimensional folds.   In this review, we summarize the AdoMet radical enzyme structures, analyze the five structures in terms of 4Fe-4S cluster, AdoMet and substrate binding, and discuss the similarities and differences between the enzymes.

## IV.B. Introduction

The AdoMet radical enzymes –also referred to as the Radical SAM enzymes (here we will use the term "AdoMet radical" to minimize confusion that may arise from alternate uses of the abbreviation "SAM") – are a newly identified enzyme superfamily[1] capable of catalyzing radical chemistry similar to that performed by the AdoCbl-dependant enzymes[2-4]. Some reactions catalyzed by AdoCbl-dependant enzymes have AdoMet radical-catalyzed counterparts, such as lysine-5,6-aminomutase (56LAM)[5,6], which has the AdoMet radical cousin lysine-2,3-aminomutase (LAM)[2,7-9] and Class II ribonucleotide reductase (RNR)[10], which is substituted under anaerobic conditions by Class III RNR (aRNR), a glycyl radical-containing enzyme activated by an AdoMet radical activase[11]. The AdoMet radical and AdoCbl-dependant enzymes have in common the 5'-dA• intermediate, a highly oxidizing and unstable radical intermediate that has never been observed, though its existence has been shown indirectly through the use of an allylic analog[12], incorporation of a radiolabel into the C5' position of the 5'-deoxyadenosine product[13] and by its covalent addition to a substrate analog to form a C-adenosylated product[14]. The AdoCbl enzymes produce 5'-dA• via homolytic cleavage of the Co-C5' bond of the corrin cofactor, whereas the AdoMet radical enzymes generate the radical by reductive cleavage of a much simpler cofactor, AdoMet[2,15]. Comparison of the two enzyme families has led to the description of AdoMet by Baker and Frey as the "poor man's adenosylcobalamin" on account of its relative simplicity and lower energetic cost of production[2], though this description of AdoMet is not fully accepted.

The AdoMet radical enzymes are attractive from an evolutionary perspective for this relative simplicity and other reasons. They require components that theoretically would have been available in the ancient world[16], and although a primordial precursor of

AdoCbl could also have existed[17], it has been theorized that AdoMet preceded AdoCbl in the prebiotic world[2]. AdoMet radical superfamily members are found spread across the kingdoms of life and catalyze a highly diverse set of reactions[1,18], again implying an ancient origin with ample time over the course of evolution to diversify. Finally, members of this family are involved in fundamental biological processes that presumably evolved early on, such as ribonucleotide reduction, cofactor biosynthesis and metabolism[3,4]. However, as typical with a topic such as evolution, there are quite a few entirely valid arguments both for and against designating the AdoMet radical enzymes as ancient[19-21]. In any case, they are themselves an important and interesting class of enzymes.

As mentioned above, the AdoMet radical enzymes all catalyze radical chemistry and are united as a superfamily by their common mechanism of radical generation[1,3,4] (Figure IV.1.a). The main hallmark of the superfamily, a conserved $CX_3CX\phi C$ motif ($\phi$ = tyrosine, phenylalanine, histidine or tryptophan), coordinates an iron sulfur cluster that is instrumental in initiation of the radical reaction. Years of excellent biochemical characterization by multiple laboratories on several family members have unambiguously shown that a typically oxygen-sensitive $[4Fe-4S]^{1+}$ cluster is the active iron-sulfur species in each of the AdoMet radical systems[22-24]. The 4Fe-4S cluster is ligated by the three cysteines of the conserved motif, leaving one iron unligated and therefore unique[25]. The second hallmark of the superfamily is a glycine-rich region involved in binding a molecule of AdoMet[1]. This bound AdoMet ligates the unique iron of the 4Fe-4S cluster[25-28] and serves as the direct source of 5'-dA•[25]. Experiments have shown clearly that this interaction with the $[4Fe-4S]^{1+}$ cluster leads to reductive cleavage of AdoMet to form the radical[29]. Though the detailed mechanism of this process is still unclear, it is

thought to occur by electron transfer from the cluster to AdoMet through the AdoMet

sulfur atom[4,27,28]. Clarification of how AdoMet activation to generate 5'-dA• is

accomplished is one area of active research in the AdoMet radical field.

Since the classification of these enzymes as a superfamily, researchers have

elucidated key details of the radical generation processes, begun characterization of new

AdoMet radical enzymes, and published the first few crystal structures of superfamily

members. This chapter will focus on the key aspects of the first five complete structures

in order to highlight the structural features of the superfamily and identify the main

elements involved in substrate binding and catalysis.


## IV.C. Summary of Reactions and Structural Studies

The AdoMet radical structures now available represent a good cross section of the

superfamily in terms of their diverse reactions and substrates[9,30-35] (Figures IV.1, IV.2).

Each structure was solved by multiple anomalous dispersion (MAD) techniques (Table

IV.1), and, as discussed below, is characterized by a partial or full $(\beta/\alpha)_8$ barrel fold (also

called the TIM barrel fold; see section IV.D.1). The structure of TYW1[34,35], an AdoMet

radical enzyme involved in biosynthesis of wyobutosine, will not be discussed here due

to the fact that its coordinates are not yet available.


### IV.C.1. PflAE

Along with LAM, PflAE is one of the best characterized of the AdoMet radical enzymes.

This enzyme is a member of the AdoMet radical activase subfamily, all of which perform

a direct hydrogen atom abstraction from a target Glycyl Radical Enzyme, forming a

catalytically essential glycyl radical (Figure IV.1.b). PflAE itself forms a glycyl radical

on $G_{734}$ of pyruvate formate-lyase (Pfl), activating it for homolytic cleavage of pyruvate to form acetyl-CoA and formate. This reaction is one of extreme importance, as under fermentative conditions, the Pfl reaction serves as the organism's sole source of acetyl-CoA.

Recent structural characterization of PflAE has yielded two models of the 4Fe-4S bound form of the enzyme, in both the substrate-free and substrate-bound states. The two structures were solved by iron-MAD techniques from two different crystal forms at 2.25 Å and 2.8 Å resolution, respectively (Table IV.1). The substrate-bound PflAE model has the 4Fe-4S cluster, AdoMet and a seven-residue peptide substrate ordered in the active site. The particular peptide used in this study, RVSGYAV, corresponds to the seven residues of the Pfl glycyl radical loop, and has been shown to be a good substrate for PflAE ($K_m$ = 0.22 mM, $V_{max}$ = 11 nmol/min·mg, compared to 1.4 μM and 54 nmol/min·mg for Pfl)[13]. The substrate-free model, which was refined to higher resolution but is less ordered, contains only the 4Fe-4S cluster, even though AdoMet was included in the crystallization conditions. Any analysis of this model should take into account disorder in several of the loops near the active site of the substrate-free PflAE model, as well as the medium resolution of the substrate-bound form. The disorder observed in the substrate-free form is likely physiologically relevant, related to conformational flexibility of parts of the enzyme in the absence of substrate (see Chapter III).

*IV.C.2. HemN*

One of the first AdoMet radical enzymes to be structurally characterized, HemN, is one of the more recently biochemically characterized members of this superfamily. It catalyzes two oxidative decarboxylations of the propionate sidechains on rings A and B

of coproporphyrinogen III to vinyl groups, yielding protoporphyrinogen IX, an important heme precursor (Figure IV.1.b). This reaction requires cleavage of two molecules of AdoMet[36,37], along with the action of an unidentified electron acceptor. Though HemN has been better characterized in recent years, full activity is still only obtained with the addition of cell free extract[36-38]. Presumably the addition of cell extract is necessary in order to provide the electron acceptor, but the fact remains that the reaction requirements are currently incompletely defined, as in several other AdoMet radical systems[3]. However, this aspect of the HemN reaction – the enzyme's inability to complete the reaction in the absence of the electron acceptor – has made it amenable to spectroscopic characterization of radical reaction intermediates[38].

The HemN structure was solved by iron-MAD techniques, refined to 2.07 Å resolution (Table IV.1), and contains the 4Fe-4S cluster, the cluster-bound AdoMet, and an additional molecule of AdoMet, termed SAM2, bound within the active site[30]. The possible role of SAM2, which could be physiologically relevant or an artifact of the crystallization conditions, is discussed in section IV.F.5. The actual substrate of this enzyme is not present in this model, and as mentioned above, the electron acceptor is also absent.

*IV.C.3. BioB*

Several AdoMet radical enzymes catalyze sulfur insertion reactions, and BioB, as well as the more recently characterized LipA, is at this point the archetype for this subfamily. BioB uses two molecules of AdoMet to insert a sulfur atom into positions C6 and C8 of dethiobiotin, forming the thiophane ring of biotin (Figure IV.1.b). The structure of BioB[31] helped resolve a controversy regarding the source of sulfur for this reaction,

substantiating the theory that a protein bound 2Fe-2S cluster is destroyed during catalysis to form biotin[3]. This was later demonstrated by incorporation of selenium into dethiobiotin to form selenobiotin[39]. It is currently unclear whether BioB is a true enzyme, because under *in vitro* conditions the protein has only ever completed a single turnover, consistent with the observed destruction of the 2Fe-2S cluster[40]. It is thought that the cluster is regenerated *in vivo* through the actions of currently unidentified enzymes[41]. Components involved in reconstitution of this cluster *in vivo* have not yet been identified, though this is an area of ongoing research.

The 3.4 Å resolution model of the BioB homodimer was solved by our laboratory using iron-MAD methods (Table IV.1). Along with HemN, it was one of the first two AdoMet radical enzymes structurally characterized, revealing the core AdoMet radical fold and confirming coordination of an iron from the 4Fe-4S cluster by AdoMet[31]. Structure based sequence alignments conducted using the BioB structure predicted the six-stranded partial TIM barrel fold of the AdoMet radical enzymes with specific motifs involved in AdoMet binding[18] (discussed in Section IV.F). In the BioB structure, the 4Fe-4S cluster, AdoMet, the dethiobiotin substrate and a catalytically relevant 2Fe-2S cluster were observed bound to the active site. The resolution of the BioB structure is relatively low, and as mentioned above, the components required for reconstitution of the 2Fe-2S cluster have not yet been identified; however, this structure comprises all of the components required for catalysis and is therefore a complete model of the BioB pre-turnover Michaelis complex. Turnover is not observed only because there is no reductant present in the crystallization conditions.

*IV.C.4. MoaA*

One of the first steps of molybdenum cofactor biosynthesis is the rearrangement of guanosine-5'-triphosphate (5'-GTP) to form precursor Z (Figure IV.1.b). Subsequent steps convert this intermediate to molybdopterin by forming the dithiolene group of molybdopterin that is responsible for coordination of molybdenum. Precursor Z differs from the intermediates of related pterin biosynthetic pathways in that C8 of the purine is retained, inserted between C2' and C3' of the ribose[42,43]. The formation of precursor Z is catalyzed by two enzymes, MoaA and MoaC, and though both structures have been solved, their specific roles during catalysis are presently unclear[32,33,44]. The identity of the actual substrate of each enzyme is also unknown, though 5'-GTP does bind to MoaA[33]. Finally, the question of whether MoaA and MoaC form a protein-protein complex is still unresolved.

Multiple structures of MoaA are available, each packing as a homodimer, with the essential 4Fe-4S cluster and an additional 4Fe-4S cluster at the C-terminus (Table IV.1). The MoaA structures solved include the "apo" (AdoMet-free) state, solved by iron-MAD, the AdoMet-bound state, solved by molecular replacement (MR) using the apo-MoaA structure as the search model, and the 5'-GTP-bound state with the AdoMet cleavage products in the AdoMet binding site, solved again by MR using the MoaA-AdoMet complex structure[32,33]. A structure of a triple mutant of MoaA has also been solved by MR with the AdoMet-bound MoaA model, in which the residues mutated (R$_{17}$/$_{266}$/$_{268}$A) are all involved in 5'-GTP binding[33]. The second 4Fe-4S cluster seems to be involved in substrate binding, ligated at a unique iron by the 5'-GTP or, in the absence of 5'-GTP, dithiothreitol (DTT). Because the specific MoaA reaction has not been fully biochemically characterized, these structures may not represent a complete description of the components required for catalysis. MoaC may indeed form a complex with MoaA.

Also, because the structure of MoaA complexed with AdoMet and 5'-GTP was obtained by soaking the MoaA-AdoMet crystals in a solution with 5'-GTP, it could be argued that crystal contacts may have prevented a conformational change required for catalysis.

### IV.C.5. LAM

Of all of the AdoMet radical enzymes, LAM is currently the best characterized[3]. It catalyzes the migration of the amino group from C2 to C3 of lysine, using pyridoxal 5'-phosphate (PLP) to bind the substrate through the typical imine linkage (Figure IV.1.b). Experiments probing the radical reaction mechanism have been conducted using substrate analogs. The LAM system has also been used to indirectly show the formation of the 5'-dA• intermediate using an AdoMet analog[12]. The reaction, essential components and mechanism of LAM have been very well characterized, and this enzyme has also been used to further investigate the details of AdoMet cleavage by this family of enzymes[45]. At this point, LAM is the prototypical AdoMet radical enzyme.

The structure of *Clostridium subterminale* 4Fe-4S-LAM in complex with AdoMet, substrate lysine and PLP was solved by iron-MAD techniques to 2.1 Å resolution[9] (Table IV.1). The model is a homotetramer, with extensive dimerization and tetramerization interfaces (Section IV.D.2.e). As was the case in the BioB crystallization conditions, turnover was prevented by omission of a reductant; therefore, this structure closely represents the Michaelis complex.

### IV.D. Overall Fold

*IV.D.1. AdoMet Radical Core*

The AdoMet radical structures reveal a common core fold responsible for radical generation to which additional protein structural elements can be added to tailor the enzyme to a particular substrate. As mentioned previously (here and elsewhere[4,9,18,30,32,46,47]), this core fold is comprised of six β/α motifs arranged in a manner that is reminiscent of a TIM barrel and is thought to be common to all of the superfamily members on the basis of structure-based sequence alignments[48]. The TIM barrel fold itself is characterized by an eight-stranded, all-parallel β sheet made up of repeating β/α motifs, curved such that the eight strands form a barrel, surrounded on the outside by the eight helices (Figure IV.3). The AdoMet radical core is similar to this fold, but contains only six of the eight β/α motifs, resulting in the formation of a "partial" or "3/4" TIM barrel (Figures IV.4 – IV.8, panels a and b, IV.9, IV.10). The helices of each β/α unit are located on one side of the sheet (the "outer" side of the barrel), while the active site is located on the opposite (or "inner") side. The curvature of the core β sheet is typically less than that of a full barrel, except in the case of BioB, which does form a complete eight-stranded barrel. In the majority of the structures, the partial TIM barrel has less curvature (appearing "splayed"), reflecting that the orientation of the β/α motifs can vary slightly depending on the remaining protein fold. Protein elements outside of the core domain are involved in imparting specificity to the individual proteins, and provide surfaces for oligomerization as well.

The lack of closure of the partial barrel results in exposure of one face of the β sheet, which forms what is here referred to as the barrel's lateral opening. This opening can be "covered" or "plugged" by protein elements outside of the AdoMet radical core, typically from the C-terminal region with some minor contributions from the N-terminus

or another molecule (Figures IV.4 – IV.8). It is within this lateral opening that the active site resides. The essential 4Fe-4S cluster and AdoMet bind via loops emanating from core $\beta$ strands. The cluster binding loop follows $\beta 1$ and harbors the canonical $CX_3CX\phi C$ motif (Figure IV.10). Binding of the 4Fe-4S cluster and AdoMet is discussed in more detail in sections IV.E and IV.F of this chapter, respectively.

Superimposition of the AdoMet radical core from each protein shows that the structure of this domain is very highly conserved (Figures IV.9). The core domain is defined here as beginning at the N-terminus of the strand that leads into the cluster-binding loop (PflAE $R_{21}$, HemN $L_{53}$, BioB $Q_{41}$, MoaA $D_{15}$, and LAM $R_{116}$) and ending at the C-terminus of the sixth strand (PflAE $P_{200}$, HemN $N_{241}$, BioB $M_{223}$, MoaA $E_{195}$, LAM $Q_{291}$), at which point the five structures begin to diverge. The first section of the core, the first strand and the cluster-binding loop, are the most varied between the five structures (Figure IV.9). The remaining protein fold shows very little variation, with strands $\beta 2$ – $\beta 6$ and the helices that follow them almost identically positioned in each structure. The helices of the core show more variation between the five structures than the strands, though one helix of note, helix $\alpha 4A$, is missing in the LAM structure (Figure IV.10). Helix $\alpha 4A$ (Figure IV.9a), which is located on the loop connecting $\beta 4$ and $\alpha 4$, is positioned along the top of the partial barrel, with the helix C-terminus pointing towards the cluster. This helix is absent in the core of LAM, but is replaced by a helix from an N-terminal domain of that protein. Finally, the structures begin to diverge at strand $\beta 6$ and the loop following it in order to appropriately accommodate substrate.

*IV.D.2. Protein elements outside of the AdoMet radical core*

The AdoMet radical core is typically located at or near the N-terminus of the molecule, while the length and sequence of the C-terminal region are highly divergent between the different AdoMet radical subfamilies. This observation leads to speculation that the C-terminal region in large part determines substrate specificity, which is a tempting conclusion but is also an oversimplification, as a significant portion of the substrate binding site (discussed in section IV.F) is provided by the core itself. While it is simpler to think of these enzymes as containing separate domains, with the substrate binding site at the interface of those domains, examination of the AdoMet radical structures (Figures IV.4 – IV.8) shows that this is not necessarily true. HemN (Figure IV.5) does appear to have two separate domains, but PflAE, BioB, MoaA and LAM each contain additional C-terminal (and N-terminal) protein elements that extend or complete the core domain. These additional elements can add oligomerization interfaces and help form the substrate and cofactor binding sites, as we will see in more detail below. A common theme in the AdoMet radical structures is embellishment or extension of the central core both by the N-terminus (as in PflAE, Figure IV.4) and by C-terminal protein elements. Strands are added to the C-terminal side of the core sheet in LAM (Figure IV.8), while in HemN (Figure IV.5) and MoaA (Figure IV.7), the C-terminus adds strands to both sides of the core and in BioB (Figure IV.6), strands are added to complete a $(\beta/\alpha)_8$ barrel. Below we look in more detail at the non-core folds of the five individual AdoMet radical structures.

### IV.D.2.a. PflAE

PflAE is by far the simplest of the five AdoMet radical structures, comprising very few structural elements outside of the core domain (Figure IV.4). An N-terminal $\beta$ strand positioned antiparallel to the main $\beta$ sheet and a C-terminus ending in a final helix and

short strand are the only additions the protein makes to the AdoMet radical core (Figure IV.4.a, b). As a result, a large surface is available for interaction with the protein substrate. The lack of additional structural elements also means there is no oligomerization surface, consistent with PflAE being a monomer.

### IV.D.2.b. HemN

HemN does contain an additional domain outside of the AdoMet radical core formed by the C-terminal region of the protein (Figure IV.5.a, b). The C-terminal region extends the sheet of the core by two antiparallel β strands and then forms a β finger motif that traverses the N-terminal side of the core sheet (orangish-yellow in Figure IV.5.c). This small motif covers the lower portion of the AdoMet radical core's lateral opening and presumably contacts substrate. The polypeptide chain then rejoins the N-terminal side of the core to further extend the β sheet on that side (orange in Figure IV.5.c). Finally, the C-terminus of HemN forms a small β/α domain that contacts the β finger motif and more fully covers the N-terminal side of the lateral opening (red in Figure IV.5.c). Forming what was termed a "trip-wire" by Layer *et al.*, the first approximately 50 residues of the HemN structure wrap around the upper part of the C-terminal β/α domain, possibly involved in substrate binding (Figure IV.5.c)[30]. The overall architecture of HemN forms a deep cleft that allows access to the lateral opening of the partial barrel fold. Though a separate C-terminal domain is formed, no oligomerization surfaces are created, as HemN is also a monomer. HemN is indeed the only enzyme that contains a full second domain possibly involved in substrate binding. This suggests that the second domain fills the need for any structural stability or other benefits that may be conferred by oligomerization.

131

*IV.D.2.c. BioB*

Unlike the other four AdoMet radical structures, BioB forms a true TIM barrel (Figure IV.6). The six-stranded AdoMet radical core is slightly more curved than the other structures and is extended at the C-terminus by two β/α motifs, which complete the barrel (Figure IV.6.a, b). The substrate binding site is located at the center of the barrel, sandwiched between the AdoMet and the 2Fe-2S cluster[31]. Several helices from the N-terminus before the start of the core form the dimerization surface, along with the non-core helices of the barrel (Figure IV.6.c). BioB is known to be a dimer[49].


*IV.D.2.d. MoaA*

The MoaA structure (Figure IV.7) is similar to HemN in that the C-terminus of the protein adds β strands to both the C-terminal and N-terminal sides of the AdoMet radical core. A set of three antiparallel β strands are appended to the C-terminal end of the core sheet (yellow and light orange in Figure IV.7.c), followed by addition of two antiparallel strands to the N-terminus (darker orange in Figure IV.7.c). The protein chain ends in a helix that is situated with its N-terminal end pointing towards the second 4Fe-4S cluster (red in Figure IV.7.c). Loops in the C-terminal region complete the substrate / second cluster binding site; specifically, those connecting the C- and N-terminal extensions and the loop before the final helix (Figure IV.7.c). The dimerization surface is formed by the final MoaA helix, the C-terminal edge of the β sheet, and the loops within and connecting these protein elements. Like PflAE, the N-terminal end of the MoaA sheet is not buried by other protein elements.

*IV.D.2.e. LAM*

The structure of LAM is a dimer of dimers, with an extensive and highly conserved dimerization surface at the bottom (N-terminal end) of the AdoMet radical partial barrel, and a C-terminal region with low overall sequence conservation that forms the tetramerization interface (Figure IV.8). The LAM monomer does have two separate domains (blue corresponds to the first domain, and green to red the second domain in Figure IV.8.c), but in this case, the second domain is at the N-terminus of the protein and does not appear to be near the substrate binding site. The LAM polypeptide chain begins with a small helical domain (blue in Figure IV.8.c) that stacks upon the top of the AdoMet radical partial barrel. This domain is likely involved in providing a buried environment for the 4Fe-4S cluster (discussed in section IV.E). The AdoMet radical core follows this domain and is then extended by two additional β strands outside of the core that form a β finger motif (orange in Figure IV.8.c) similar to that seen in the HemN structure (yellow in Figure IV.5.c). The remainder of the LAM C-terminus is involved in a domain swap essential to oligomerization. A long loop after the main LAM domain extends towards the surface of a second molecule, where the protein forms a three-stranded antiparallel β sheet that appears to form the majority of the tetramerization interface (labeled "A" and colored orange in Figure IV.8.c; part of a LAM chain that forms the tetramer is shown in peach in Figure IV.8.a,b). Finally, the C-terminus (red in Figure IV.8.c) wraps around the tetramer, providing more extensive interactions with the other molecules (the second molecule is colored dark grey and the remaining two molecules of the tetramer are colored light grey in Figure IV.8.c).

The substrate binding site of LAM is created by the loop preceding the core (blue in Figure IV.8.c), the lateral opening within the AdoMet radical core itself and the β

finger motif following the core (orange in Figure IV.8.c). The remainder of the binding

site is formed by the loop leading from the core into the β finger motif of the second

LAM molecule (colored dark grey in Figure IV.8.c). Quite a large amount of the C-

terminus of the LAM polypeptide chain facilitates oligomerization, though (as in BioB

and MoaA), there does not appear to be a large substrate binding domain *per se*. Rather,

the substrate binding site is formed by the addition of several structural elements and

loops outside of the AdoMet radical core. Finally, the presence of the second LAM

polypeptide deeply buries each LAM active site at the N-terminal side of the AdoMet

radical core.


## IV.E. The FeS cluster

*IV.E.1. Location of the 4Fe-4S cluster binding site*

The most well known feature of the AdoMet radical enzymes is their use of three

cysteines from a completely conserved $CX_3CX\phi C$ motif to bind an anaerobically stable,

site-differentiated 4Fe-4S cluster[25]. The unique iron of the cluster is coordinated by the

methionyl moiety of a bound molecule of AdoMet[25-28].

The 4Fe-4S cluster is bound at the C-terminal end of the AdoMet radical partial

β-barrel core domain, located above the lateral opening discussed above (Figures IV.4 –

IV.9). The $CX_3CX\phi C$ motif, also as described above, resides on a loop (termed the

cluster-binding loop) following the first β strand of the AdoMet radical core. The length

of this particular loop varies from enzyme to enzyme, depending on the spacing of the

β/α motifs surrounding it. PflAE and LAM have cluster-binding loops of similar length

and orientation, but those of the other three structurally related enzymes differ widely

(Figure IV.9). The cluster-binding loop length varies from approximately 20 residues in PflAE and HemN to 30 residues in BioB. The loop winds around the cluster from "behind" (closer to the core helices) to the front of the cluster, towards the lateral opening of the AdoMet radical core. Residues following the $CX_3CX\phi C$ motif are therefore located near the substrate, as discussed in section IV.G.2 below. This loop may also provide part of the interaction surface for binding of the physiological reductant (see section IV.H).

*IV.E.2. The environment surrounding the cluster*

The nature of the chemistry catalyzed by the AdoMet radical enzymes suggests that the 4Fe-4S cluster should be fully protected from solvent, and this seclusion is for the most part observed in all of the structures. Solvent exposure is observed in MoaA, but the form of the enzyme crystallized may not be the full Michaelis complex due to the absence of MoaC and therefore not a completely representative picture of catalysis. Various degrees of solvent exposure and/or burial of the cluster are observed in each enzyme. For example, the 4Fe-4S cluster of the substrate-free MoaA is solvent exposed, while that of LAM is completely buried. The cluster is typically located approximately 7-10 Å from the nearest protein surface, which in all five structures except for MoaA is formed opposite the lateral opening by the cluster-binding loop. Because the 4Fe-4S cluster binds at the C-termini of the core β strands but inside the ring of helices, it is buried by the loops of the C-terminal end of the partial barrel. Helix α4A, which is structurally conserved in all of the AdoMet radical enzymes except LAM, plays a large part of burial of the cluster. In LAM, this helix is not part of AdoMet radical core; rather, it is replaced by a helix contributed by the N-terminal helical domain of this enzyme,

which suggests a vital role for helix α4A. Another important element involved in burying

the 4Fe-4S cluster is the aromatic residue of the CX₃CXφC motif, which is conserved in

all of the AdoMet radical enzymes except the aRNR activase. Finally, several of the

enzymes, PflAE and HemN in particular, have additional long loops immediately

following the AdoMet radical core that fold over and presumably help bury the cluster.


*IV.E.3. Interactions between the 4Fe-4S cluster and AdoMet*

The crystal structures of AdoMet radical enzymes confirmed the results of several

spectroscopic studies that showed that AdoMet binds directly to the 4Fe-4S cluster[25-28].

Ligation to the cluster presumably helps to secure AdoMet in the binding site and

properly positions it for radical generation. Coordination of the unique iron by the amino

nitrogen and carboxylate group has also been shown spectroscopically for PflAE and

LAM, and several distance estimates were obtained, such as the AdoMet carbonyl carbon

to iron and methyl carbon of AdoMet to iron distances (Table IV.1). Again, the

structures have been, for the most part, consistent with the spectroscopic results (Table

IV.1). Differences between enzymes are observed, which may be a result of variations in

constraints used during refinement, the variety of resolutions obtained for the structures,

or real deviations in the coordination geometries between enzyme family members.

Though there is some variation between distances observed for the direct interactions

between the unique iron and the AdoMet methionyl moiety, the biggest differences are in

the distances from the cluster to the AdoMet sulfur atom, which range from 3.1 Å (LAM)

to 4.0 Å (BioB) to the closest iron and from 3.4 Å (MoaA) to 4.2 Å (BioB) to the nearest

sulfide, and from the methyl carbon atoms to the iron, with a range of 3.5 Å (PflAE) – 5.4

Å (BioB). Despite these variations, superimposition of the coordinates of the 4Fe-4S

cluster and AdoMet from each of the structures (Figure IV.11) shows that the sulfur atom

occupies a nearly identical position with respect to the cluster; therefore, the distance

differences observed with respect to the AdoMet sulfur atom likely result from the

combined discrepancies in refinement parameters of the cluster, AdoMet and cluster-

AdoMet ligation. However, the methyl group position does vary significantly between

the structures. This may result, again, from discrepancies in refinement parameters or,

more likely, from differences in the overall AdoMet positioning by the individual

enzymes. This, along with other differences in the AdoMet conformation bound by the

enzymes, may reflect their ability to tailor the active site to the specific reaction being

catalyzed. In the following section, the AdoMet binding site is examined in more detail

to understand more fully how the enzymes ligate AdoMet in order to produce 5'dA•.


**IV.F. AdoMet binding**

*IV.F.1. AdoMet conformation*

The AdoMet conformation observed bound to the AdoMet radical enzymes is similar in

each structure (Figure IV.11). The molecule always adopts an anti conformation at the

glycosidic bond in these enzymes, but varies in terms of the puckering of ribose (3'-endo

in the majority of the enzymes, with PflAE and MoaA as the exceptions). In terms of the

position of the methionine with respect to the 5'-dA moeity, AdoMet can assume a

variety of conformations, either "folded" or "extended" based on the torsional angle at

atoms O4'-C4'-C5'-SB (referred to as the $\psi$ torsional angle)[50]. The lowest energy

extended conformation of AdoMet has a $\psi$ torsional angle of ~180°, compared to <90°

for folded conformations. In contrast to the majority of protein structures with bound

AdoMet, all of the structures of AdoMet radical enzymes bind AdoMet in the folded

conformation, with the sulfonium located close to O4' of the ribose ring. HemN is the outlier of the group on this point, binding AdoMet with a $\psi$ torsional angle of -112.80°, compared to -48.75° (BioB), -73.96° (MoaA), -60.42° (LAM), and -64.29° (PflAE). The implications of the folded state of AdoMet observed in these enzymes are still being established, though it has been observed that when AdoMet is bound with the purpose of serving a reactive role (transfer of one of the groups bound to the sulfonium), it adopts an extended conformation, and when bound as an activator or substrate, as in MetJ and AdoMet decarboxylase, it binds without a preference for the extended $\psi$ torsional angles. The overall conformation AdoMet adopts when bound to an AdoMet radical enzyme is likely extremely important to enable electron transfer from the cluster. It is notable that although the conformations of AdoMet observed amongst the AdoMet radical enzymes do differ somewhat, superposition of the 4Fe-4S cluster and AdoMet coordinates from each enzyme reveals that all of the enzymes place the AdoMet sulfur atom in a highly similar position with respect to the 4Fe-4S cluster (Figure IV.11). This observation is consistent with the theory that electron transfer occurs to the AdoMet sulfur atom.

The anchoring of AdoMet by an iron-sulfur cluster via the methionyl amino and carboxyl atoms had, of course, never been observed in a protein before structural characterization of an AdoMet radical enzyme, all of which employ a unique structural motif for binding AdoMet.

*IV.F.2. General properties of the AdoMet binding site*

In all of the AdoMet radical enzyme structures[9,30-33,51], AdoMet is observed bound across the top (C-terminal end) of the partial or full TIM barrel in a binding site made up of residues from each of the core $\beta$ strands (Figures IV.10, IV.12). The binding site is

138

hydrophilic and provides residues for several specific interactions with AdoMet (see below). Although little conservation is observed across the superfamily, many of the residues involved in binding AdoMet and/or substrate are highly to fully conserved within individual AdoMet radical subfamilies, such as the BioB protein subfamily.

The nature of the chemistry performed by the AdoMet radical enzymes suggests that the active site would require burial within a protected protein environment during catalysis to prevent quenching of the radical. This sequestration of the active site from solvent is indeed observed in the BioB, LAM and substrate-bound PflAE structures[9,31,51]. Formation of the full PflAE-Pfl complex, as opposed to the smaller peptidic substrate observed bound in the crystal structure[51], will certainly result in even more extensive burial of the active site. The HemN and MoaA structures do not provide a complete view of the substrate-bound forms of those enzymes, and as one would expect, their active sites are solvent exposed[30,32,33]. Though several structures of MoaA are available, there is no pre-catalysis substrate-bound form for analysis of the buriedness of the AdoMet binding site at that stage of turnover (Table IV.1). Additionally, the MoaA reaction requires presence of another protein, MoaC, whose function during catalysis is unknown[44]. Further investigation of these two enzymes, and structural characterization of more of the AdoMet radical superfamily members, will clarify this aspect of catalysis.

### IV.F.3. Overall description of the AdoMet binding site

Low sequence conservation within the AdoMet radical superfamily may arise in part from the location of the AdoMet binding site in these proteins. Because AdoMet binds across the top of the β barrel-like sheet, the residues of the binding site originate from loops following each β strand, and are therefore dispersed throughout the primary

sequence (Figure IV.10). This dispersion hinders identification of a specific AdoMet

binding motif. Low sequence similarities notwithstanding, comparison of the five

currently available AdoMet radical structures shows that they all employ remarkably

similar modes of interaction with AdoMet. Examination of the dendrogram visualization

of the AdoMet radical core domains constructed by Sofia *et al.*[1] demonstrates that the

AdoMet radical structures we now have are distantly related within the superfamily and

represent a good cross-section of the sequence space covered by the individual members.

Therefore, the structural comparison conducted here allows us to provide a general

description of AdoMet binding in this superfamily that is likely applicable to all of the

members (Figure IV.10).


### IV.F.3.a. The AdoMet methionyl moiety

As discussed previously, the methionyl carboxylate and amino groups of AdoMet ligate

the unique iron of the AdoMet radical 4Fe-4S cluster in order to help anchor and orient

the molecule for electron transfer, and possibly alter the redox potential of the 4Fe-4S

cluster and/or AdoMet. In each structure, the amino group is observed binding close to

the C-terminus of strand $\beta2$, making hydrogen bonds with backbone atoms of the "GGE

motif" – also termed the "glycine rich region" – located on that strand (Figure IV.12;

PflAE $G_{77}$, $G_{78}$, $E_{79}$; HemN $G_{112}$, $G_{113}$, $T_{114}$; BioB $A_{100}$, $A_{101}$, $W_{102}$; MoaA $G_{74}$, $G_{75}$, $E_{76}$;

LAM $G_{170}$, $G_{171}$, $D_{172}$). In some cases the sidechain of the residue at position 3 of this

motif is within hydrogen bonding distance of the AdoMet amino group (Figure IV.12).

The hydrogen bonds provided by this motif are the only interactions observed to the

amino group, and ensure the correct orientation of this part of AdoMet in the active site,

as well as proper coordination of the unique iron by the methionyl moiety. In contrast,

much more variation is observed with respect to contacts with the AdoMet carboxylate by the AdoMet radical enzymes. In fact, it is this particular interaction with AdoMet that varies the most between the individual enzymes in terms of location within the primary sequence of the residues involved, with residues originating from β3, the loop following β3, α3 and β4 (Figures IV.10, IV.12). The majority of the enzymes use a conserved arginine, lysine or histidine (PflAE $K_{131}$; HemN $R_{184}$; BioB $R_{173}$; LAM $H_{230}$) to bind and orient the carboxylate, while the MoaA binding site provides threonine and serine hydroxyls for hydrogen bonding ($T_{102}$, $S_{128}$). Positioning of this portion of AdoMet is further fine tuned by interactions with backbone atoms (PflAE $T_{105}$; HemN $V_{146}$ and $G_{170}$; BioB $G_{132}$; MoaA $N_{104}$), additional sidechain atoms (PflAE $D_{104}$ and $D_{129}$) or, as in LAM, a tightly bound water molecule. These variations in binding likely yield the different conformations of the AdoMet methionine in the five structures discussed here (Figure IV.11), ranging from the "compact" conformation of the methionine moiety observed in the LAM structure to the "elongated" one bound to BioB.

*IV.F.3.b. The AdoMet ribose*

Some variation in the AdoMet radical enzymes is also evident in the details of their interactions with the AdoMet ribose. Residues responsible for interacting with ribose originate mainly from β4 (Figures IV.10, IV.12) with a few exceptions (those exceptions include HemN $D_{209}$ and LAM $Q_{258}$, both from β5). The ribose hydroxyls can interact with charged or polar protein sidechains directly (PflAE $D_{129}$; HemN $Q_{172}$, $D_{209}$; BioB $N_{153}$, $D_{155}$; MoaA $S_{126}$; LAM $Q_{258}$) or via waters bound in the active site (MoaA $D_{128}$). In some of the structures, the hydroxyl of the $CX_3CX\phi C$ motif tyrosine (PflAE $Y_{35}$; BioB

$Y_{59}$; MoaA $Y_{30}$) is within 4 Å of a ribose hydroxyl, though it is unlikely that this residue plays a major role in ribose binding because the hydroxyl group is not fully conserved and this residue is more important in terms of adenine binding (see below) and protection of the cluster from solvent. Finally, with the exception of BioB, each AdoMet radical enzyme has a highly conserved charged or polar residue within strand $\beta5$ that interacts with or is located very near the AdoMet ribose and/or substrate (PflAE $R_{166}$; HemN $D_{209}$; MoaA $N_{165}$; LAM $Q_{258}$). Within each enzyme subfamily, that residue is positioned close to the ribose hydroxyls and C5′ of AdoMet as well as the substrate. Moreover, many AdoMet radical enzyme subfamily members, not just those with known structure, appear to have a similarly conserved residue at the same place in $\beta5$ (data not shown). The high conservation and position of that residue in the structures (PflAE $R_{166}$; HemN $D_{209}$; MoaA $N_{165}$; LAM $Q_{258}$) suggests an important role during catalysis such as in the orientation of 5′dA• and/or substrate for proper hydrogen atom abstraction.

### IV.F.3.c. The AdoMet adenine moiety

When bound to an AdoMet radical enzyme, the adenine moiety rests against 3-5 hydrophobic residue sidechains and is specifically recognized by 3-4 hydrogen bonds made to protein backbone atoms. This portion of the binding site is made up mainly by residues from $\beta$ strands 5 and 6, with contributions from the $CX_3CX\phi C$ motif of the cluster-binding loop (Figure IV.10). The interactions with adenine are maintained in the superfamily and are more easily predicted than other interactions, though enzyme evolution has resulted in a variety of compensatory mutations, particularly at those residues involved in protein backbone to AdoMet hydrogen bonds. Hydrophobic interactions are invariably provided by three specific regions of the primary sequences

(Figures IV.10, IV.12): (1) the conserved aromatic residue of the $CX_3CX\phi C$ motif ($Y_{35}$ in PflAE; $F_{68}$ in HemN; $Y_{59}$ in BioB; $Y_{30}$ of MoaA; $H_{131}$ of LAM), (2) a residue from $\beta 5$, located at position 2 of a conserved motif utilized in part for adenine binding, termed the GxIxGxxE motif ($V_{168}$ of PflAE; $I_{211}$ of HemN; $I_{192}$ of BioB; $V_{167}$ of MoaA; $V_{260}$ of LAM) and (3) one or two residues located at the end of $\beta 6$ that form a conserved structural motif ($L_{199}$, $H_{202}$ of PflAE; $F_{240}$, $Y_{242}$ of HemN; $V_{225}$ of BioB; $I_{194}$, $F_{196}$ of MoaA; $Y_{290}$ of LAM). Hydrogen bonds are made to atoms N1 and N6 of AdoMet by backbone atoms. These backbone hydrogen bonds are made by the aromatic residue of the $CX_3CX\phi C$ motif and by a residue from the adenine-binding structural motif of $\beta 6$ ($H_{202}$ of PflAE; $A_{243}$ of HemN; $V_{225}$ of BioB; $M_{197}$ of MoaA; $D_{293}$ of LAM). Lastly, N7 of AdoMet is in some cases hydrogen bonded by a backbone atom from the residue immediately following the $CX_3CX\phi C$ motif (PflAE $Y_{35}$, $H_{37}$; HemN $G_{70}$).

*IV.F.3.d. AdoMet binding motifs in the AdoMet radical superfamily*

The solution of an AdoMet radical structure was crucial for identification of a preserved AdoMet-binding protein fold and common AdoMet binding motifs in this superfamily. The current availability of several structures has allowed identification of a conserved structural motif in $\beta 6$, as well as clarification of one in $\beta 5$ that had been previously identified (GxIxGxxE), both of which interact with the AdoMet adenine (Figure IV.10). The very low sequence homology between members of the AdoMet radical superfamily prevents unambiguous naming of the motifs in a manner incorporating residue names. This may lead to some confusion, as the motif names do not always correspond to the sequence of each enzyme. However, the AdoMet binding motifs will be referred to in

143

this discussion by the names they have already been given, and this section will attempt

to clarify the function of each motif.

The $CX_3CX\phi C$ motif of the AdoMet radical superfamily functions mainly in

ligation of the cluster (see above). In addition, the aromatic residue present within this

motif in every enzyme except aRNR activase always provides hydrophobic interactions

to the adenine binding site and hydrogen bonds N6 of adenine via the backbone oxygen.

The "GGE motif" or "glycine-rich region" observed at the end of strand $\beta2$ invariably

binds the amino group of AdoMet, aiding in proper ligation of the cluster's unique iron.

This motif also forms the wall of this side of the active site and maintains the structure of

the loop region after $\beta2$.

The GxIxGxxE motif ($\beta5$) and the $\beta6$ structural motif of the AdoMet radical

superfamily form most of the AdoMet adenine binding site. The GxIxGxxE motif,

named in reference to the BioB sequence, provides hydrophobic interactions to the

adenine portion of the binding site at its second conserved position (Figures IV.10 and

IV.13; PflAE $V_{168}$; HemN $I_{211}$; $I_{192}$ in BioB; MoaA $V_{167}$; LAM $V_{260}$). Also, through one

or more conserved sidechain-to-backbone interactions, this motif helps to preserve the

structure of the AdoMet binding site as well (Figure IV.13). For example, the interaction

observed in BioB from $E_{197}$ (position 5) to the backbone nitrogen of $G_{194}$ (position 3) is

found in each of the five structures, though the residue at position 5 in the other structures

also interacts with the carboxyl backbone atom of the residue at position 2 (Figure

IV.13). In addition, PflAE, MoaA and LAM share a conserved sidechain-to-backbone

interaction from position 6 (PflAE $D_{175}$; MoaA $D_{174}$; LAM $D_{267}$) to the backbone amino

group of the residue in position 4. Sequence alignments suggest that this particular

interaction may be conserved in several other AdoMet radical enzymes, such as PylB,

NifB, BssD and possibly ThiH (Figure IV.10). Finally, position 1 of the GxIxGxxE motif is more difficult to characterize, as it is highly conserved but its identity differs in the individual AdoMet radical subfamilies. As discussed in section IV.F.3.b above, the positioning of this residue near the AdoMet ribose and/or substrate in most of the known structures (Figure IV.13), along with its high conservation within each subfamily, implicates it as an important residue.

The β6 structural motif, which does not have a consensus sequence, provides hydrophobic interactions to the adenine from the residue in position 1 (with the exception of BioB, which uses this residue to interact with substrate). The residue in the second position of this motif hydrogen bonds to the AdoMet adenine N1 and N6 atoms via its backbone amino nitrogen and carboxyl oxygen atoms (Figure IV.12).

*IV.F.4. Deviations and variations in AdoMet binding between the subfamilies*

The major differences between these enzymes are readily apparent from the sequence alignment in Figure IV.10 and Figure IV.12. The modes of interaction adopted by these enzymes to bind the AdoMet ribose and carboxylate groups specifically vary widely. As mentioned above, AdoMet carboxylate binding by these enzymes governs the position of the methionyl moiety, resulting in the elongated or compact conformations observed (Figure IV.11). For example, $R_{173}$ of BioB is located at a distance further away from the binding site, resulting in a more elongated methionyl conformation, while the location of LAM $H_{230}$ in the center of the binding site causes the methionyl group to bend inwards (Figure IV.12.c, e). Note that the methionyl conformation is independent of the overall AdoMet folded or extended conformation, which is a function of the ψ torsional angle. In addition, the hydrogen bonds formed to the AdoMet amino group by motif 2 vary

145

between the enzymes. The specific residue of the motif providing backbone atoms for hydrogen bonding can vary, as well as involvement of the last residue's sidechain. This does not effect the position of the AdoMet amino group. Comparison of the ribose binding interactions are less straightforward, with water mediating the contacts in MoaA and LAM, involvement of backbone atoms in LAM, and the use of both charged and polar residues in HemN and BioB. Beyond use of residues from mainly strand β4, and the presence of a Glu or Asp nearby (with the exception of LAM), the AdoMet radical enzymes share little similarity in this region of the AdoMet binding site.

Several minor differences are observed between the five structures in their mode of interaction with the adenine of AdoMet, including the number and identity of the contacting hydrophobic residues. The details of the hydrogen bonding to N6 and N7 of the adenine ring also differ between the five structures. These differences are likely a result of experimental differences such as presence or absence of substrate, the resolution at which the structure was refined, or effects of the crystallization conditions. Overall, the AdoMet adenine binding sites are more or less the same.


*IV.F.5. Implications of the AdoMet binding site architecture on function and reactivity*

The parts of the AdoMet binding site that are conserved across the AdoMet radical enzymes are limited to those interactions that are made to the adenine and methionyl amino group. This conservation results in a similar placement of those parts of AdoMet in all of the structures (Figure IV.11) and variations in the other groups, with the exception of the sulfur atom. The sulfur atom in each structure is located in roughly the same position, presumably to allow and optimize electron transfer to AdoMet from the cluster[52,53]. The placement of the ribose and carboxyl portions of AdoMet differ more

significantly in each enzyme than do the AdoMet adenine and methionyl amino groups, which likely allows each enzyme to tailor the reaction to its unique substrate. Conformational flexibility of AdoMet around its C5' atom presumably allows the individual enzymes to more carefully control the position of the molecule to ensure proper hydrogen atom abstraction from substrate upon generation of 5'-dA•.

An interesting question concerning the AdoMet binding site of these enzymes is whether any specific structural features accompany use by some AdoMet radical enzymes, LAM and SPL in particular, of AdoMet as a cofactor instead of as a cosubstrate. Comparison of the five AdoMet binding sites has not identified any obvious structural features that are related to this question. AdoMet does not appear to have more extensive interactions in LAM as compared to the other enzymes; indeed, the HemN and MoaA binding sites seem to provide more numerous contacts to AdoMet. $R_{134}$ of LAM does appear capable of physically restricting movement of AdoMet out of the binding site by folding across the methionyl moiety of the cofactor to contact substrate (Figure IV.12.e) and spectroscopic studies have shown that Met remains bound to the cluster after cleavage of AdoMet during turnover[52]. However, $R_{134}$ is not conserved in SPL, and the crystal structure of MoaA with 5'-GTP shows the AdoMet cleavage products 5'-dA and Met bound to the enzyme, demonstrating that the simple ability to retain the AdoMet cleavage products does not affect how the enzyme uses AdoMet. Other factors such as the requirements of the specific chemical reaction being catalyzed must govern this difference in AdoMet usage. For example, in the LAM reaction, there is no net change in oxidation state of the substrate, whereas in the PflAE reaction, a proton and electron are transferred irreversibly from substrate to 5'-dA•, and in the HemN reaction, two of the propionate side chains of coproporphyrinogen III are oxidized by two electrons each, thus

requiring an additional electron acceptor[36,38,54]. Structural characterization of SPL or identification and characterization of another AdoMet radical enzyme that uses AdoMet as a cofactor will be necessary to clarify any AdoMet binding site characteristics, if any exist, involved in this aspect of catalysis.

An additional question remaining with regards to AdoMet binding is how some enzymes accomplish the cleavage of two molecules of AdoMet while others use only one. Although initially controversial[55,56], studies conducted on BioB[57-59], LipA[60] and HemN[37] that quantitate and correlate AdoMet cleavage and product formation have shown that these three enzymes require cleavage of two molecules of AdoMet for every one molecule of product formed. It now remains to be shown how and where these enzymes bind the second AdoMet and in what order the substrates / molecules of AdoMet bind, including whether the enzymes actually bind two molecules of AdoMet simultaneously during catalysis. These enzyme systems, however, are wrought with many complicating factors that make even a complete description of the requirements for catalysis difficult. BioB and LipA, for example, have still only been shown to turn over once *in vitro*; in fact, 50 $\mu$M LipA only produces 18 $\mu$M of the lipoyl cofactor product[60]. Formation of the protoporphyrinogen IX product of HemN requires addition of crude cell extract to assay mixtures[36,38], presumably to include the reaction's as of yet unidentified electron acceptor. As a result, the complete requirements for full HemN activity are undefined. Clearly, our understanding of the biochemistry of these three enzymes is still in its infancy.

There are three possible ways to explain the observed stoichiometry of AdoMet cleavage in BioB, LipA and HemN. First, as proposed for LipA, each individual protein monomer could carry out AdoMet cleavage once, releasing a stable intermediate and the

148

AdoMet cleavage products[60]. However, both sulfur atoms incorporated into LipA's

lipoate product are likely derived from the cluster of the same LipA polypeptide[61].

Further, there is no evidence that BioB releases a stable intermediate and even though 9-

mercaptodethiobiotin can be turned over into biotin by BioB[62], the apparent $K_m$ for this

molecule is much higher than that of dethiobiotin[63], and use of 9-mercaptodethiobiotin

does not increase the turnover numbers of the enzyme[63]. A second possibility is that each

enzyme binds and cleaves one molecule of AdoMet at a time, releasing the cleavage

products and then binding a second AdoMet molecule, all the while retaining the

substrate intermediate. This possibility is the most consistent with what we know of

BioB, since there does not appear to be room in the BioB active site for a second AdoMet

molecule to bind[31], only minor readjustments of the BioB structure would be needed to

allow release of the AdoMet cleavage products[31] and the intermediate formed could stay

attached to the 2Fe-2S cluster[59]. The final possibility, suggested by the observation that

two AdoMet molecules bind within the active site in the HemN structure[30], is that there

are two distinct AdoMet binding sites on the enzyme. This second AdoMet molecule

(termed SAM2) is either physiologically relevant or an artifact of the crystallization

conditions, binding where substrate would normally bind. Layer et al. hold that SAM2 is

involved in catalysis based on their modeling of the coproporphyrinogen III substrate in

the active site, both with and without presence of SAM2[30]. In their analysis, the

propionate sidechains could be better accommodated in the presence of SAM2[30].

Additionally, mutation of the residues contacting the structure's SAM2 affects AdoMet

cleavage[37]. Contrary to the conclusions Layer et al. draw, however, the decreased

AdoMet cleavage they observe is consistent with those residues forming either a second

AdoMet site or the substrate binding site.

Several factors strongly suggest an artifactual explanation of SAM2 binding. First, each of the four other structures shows that substrate binds in a very specific location with respect to AdoMet, with the substrate atom from which hydrogen atom abstraction occurs approximately 4Å from the C5' atom of AdoMet (section IV.G.2). SAM2 of HemN is observed binding in this location (see section IV.G.2.b). Second, there is no SAM2 binding site observed on BioB[31], though BioB is known to utilize two molecules of AdoMet during catalysis[57-59]. Third, the nature of the chemistry catalyzed would require close control of the orientation of each AdoMet and substrate molecule involved in the reaction, and the SAM2 site does not indicate the same high level of conformational control as exercised over the cluster-bound AdoMet. Although SAM2 is recognized reasonably well by HemN through six possible hydrogen bonds mainly to the ribose hydroxyls and hydrophobic interactions with three HemN residues, there are fewer direct contacts to SAM2 than to the first AdoMet, suggesting that recognition of SAM2 is less specific; indeed, the methionyl moiety of SAM2 is partially disordered[30] (see section IV.G.2.b). If SAM2 binding was physiologically relevant, one would expect to see a tighter conformational control, as is wielded over the cluster-bound AdoMet (Figures IV.12.a-e) and the substrates in the other structures (discussed in section IV.G.1). Finally, the mechanisms Layer *et al.* propose for cleavage of the second AdoMet molecule in the SAM2 site are troubling[30,38]. First, they propose that a conformational change in SAM2 occurs upon substrate binding, which would bring the sulfonium of SAM2 into sufficiently close proximity to the cluster-bound AdoMet to allow electron transfer to occur from the 4Fe-4S cluster to SAM2 via the cluster-bound AdoMet. Their second proposal is that AdoMet cleavage occurs as is generally accepted to catalyze the first substrate decarboxylation, followed by injection of the electron released by this

process into SAM2, resulting in formation of the second 5'-dA•. Because the reduction potential of a typical sulfonium ion is so high, achievement of such a difficult mechanistic step by either of these two mechanisms – electron transfer through AdoMet or reductive cleavage of AdoMet via a decarboxylation reaction – is unprecedented and seems unlikely, especially when involvement of a presumably more suitable electron acceptor is already expected. On the other hand, model systems have confirmed that electron transfer to and subsequent cleavage of AdoMet can be mediated by iron-sulfur clusters[64,65]. The generally accepted theory for production of 5'-dA• from AdoMet therefore centers on inner-sphere electron transfer to the AdoMet sulfonium enabled by close proximity with the 4Fe-4S cluster[4,27,28], and experiments aimed at clarifying this mechanism are ongoing. Since components of the HemN reaction such as the physiological electron acceptor have not been identified, these questions await further biochemical and structural studies for clarification.

*IV.F.6. Other known AdoMet-binding protein folds*

Currently, the PDB (at the RCSB, http://www.rcsb.org/pdb) contains 295 total protein structures of known AdoMet-binding proteins, 158 of which are unique (<95% identical by sequence), whose coordinates include AdoMet (88 total, 72 unique) or S-adenosylhomocysteine (AdoHCy – 207 total, 105 unique). The large majority of these proteins are methyltransferases (MTases). As discussed by Kozbial and Mushegian[19], there are 15 distinct folds capable of binding AdoMet or AdoHCy, which can be divided into the following 4 categories: (1) Rossmann-like, (2) β barrel-like, (3) "Double-β" and (4) folds derived from other ligand-binding domains.

The most common fold utilized to bind AdoMet is the Rossmann-like domain. The typical MTase AdoMet-binding fold is a 7-stranded β sheet, adopting a mainly parallel orientation with one strand, typically β7, oriented antiparallel to the others (Figure IV.14). AdoMet binds at the C-terminal loops following strands β1, β2, β3 and β4. The Rossmann-like AdoMet-binding fold includes most of the MTases, such as catechol-O-MTase (COMTase)[66] and the DNA:m5C MTases[67], as well as some enzymes that bind AdoMet but are not MTases, like spermidine synthase[68], the bacterial fluorinating enzyme[69] and cyclopropane fatty acid synthase[70]. Additionally, several classes of MTases employ a similar fold, such as the SPOUT MTases, thus named for the first identified enzymes of this class, **SpoU** and TrmD MTases[71,72]. The Rossmann-like domain is similar to the β barrel-like domain observed in the AdoMet radical enzymes in that the loops near the C-terminal ends of the β-sheet structure are responsible for binding AdoMet. Also, these fold types are both commonly found in the context of larger protein chains that contain both an AdoMet-binding region and a variable substrate-binding region. Low sequence conservation like that observed between members of the AdoMet radical superfamily is also observed between the Rossmann-like AdoMet binding proteins.

Another less common fold able to bind AdoMet is the double-β fold, in which AdoMet binds between two β sheets. Examples of the double-β fold include the SET domain (named because the first proteins identified with this conserved domain were the Su(var)3–9, E(Z) and Trithorax chromatin remodeling proteins[19,73,74]), such as that found in PrmA MTase [75] (Figure IV.15.a), a class of MTases that transfer a methyl group to a nuclear protein lysine residue, and AdoMet synthetase[76] (Figure IV.15.b). Still other

folds were generated by recruitment of broad-specificity ligand-binding folds to bind AdoMet, such as in the case of ACC synthase (AdoMet methylthioadenosine lyase, a member of a diverse PLP-dependant transferase family)[77] and the MetJ repressor (Figure IV.15.c, a ribbon-helix-helix DNA-binding protein that uses AdoMet as a co-repressor)[78]. Finally, several enzymes, such as the methionine synthase (MetH) reactivation domain (Figure IV.15.d)[79], do not fit easily into any of these categories. The usual "cup-like" shape of the MetH reactivation domain probably evolved to specifically recognize the AdoCbl-binding domain of MetH. The AdoMet radical enzymes clearly form their own unique class of AdoMet-binding proteins with little similarity to the other known AdoMet-binding folds.

## IV.G. Substrate binding to AdoMet radical enzymes

### IV.G.1. Overall description of substrate binding site

The tight conformational control over AdoMet implied by the structures of the AdoMet radical enzymes is likely also necessary over the substrates for proper orientation and, in particular, the atom from which hydrogen atom abstraction occurs. Superposition of the five structures of the AdoMet radical core shows that substrate / SAM2 (from HemN) binds in a very similar site on all of the enzymes (Figure IV.16). The AdoMet radical substrate binding site is located within the lateral opening of the core's partial TIM barrel, which corresponds to the center of the full TIM barrel of BioB. As discussed above (section IV.D.2), the substrate binding site is formed between the core region and structural motifs located outside the core mainly at the C-terminus (Figure IV.10, Figure IV.17). As would be expected based on the location of the substrate binding sites, those core residues involved in substrate binding are strictly from the core β strands, with only

three exceptions (PflAE $H_{37}$, $N_{38}$; LAM $R_{134}$, which all immediately follow the

$CX_3CX\phi C$ motif, Figure IV.10).

A main function of the AdoMet radical substrate binding sites is likely to be

conformational control over the substrate in order to facilitate proper hydrogen atom

abstraction. There are typically several specific interactions made by the AdoMet radical

enzyme to its substrate (discussed in section IV.G.2). In addition to protein sidechain

interactions, some of the enzymes use a separate cofactor to help anchor substrate, such

as the PLP of LAM and the second 4Fe-4S cluster of MoaA. The combined interactions

between protein, cofactors and substrate result in a similar position of the substrate

abstraction point (Figure IV.18). However, the role of additional cofactors is not always

for substrate binding. While the PLP of LAM and the 4Fe-4S cluster of MoaA are

presumably used for substrate binding, the 2Fe-2S cluster of BioB is believed to serve as

a source of sulfur during catalysis[59]. The positions of the additional cofactors differ

accordingly between the enzymes (Figure IV.16), reflecting the plasticity of the AdoMet

radical fold in its ability to accommodate different substrates. In the following sections,

we will examine the substrates and cofactors of the AdoMet radical enzymes of known

structure, their individual substrate binding sites, and any identified substrate-binding

motifs.


*IV.G.2. The substrate binding sites of AdoMet radical enzymes*

  *IV.G.2.a. PflAE*

Pfl, a homodimeric enzyme of 85 kDa per monomer, is the actual substrate of PflAE. In

order to simplify the crystallization experiment, a small peptide corresponding to the

glycine residue of interest and the loop upon which it resides was used to obtain crystals

of PflAE (Section IV.C.1). In experiments conducted by Knappe's laboratory, this particular peptide (RVSGYAV) was shown to bind PflAE and stimulate its AdoMet cleavage activity, proving that the peptide is a good substrate for this enzyme[13]. The seven-residue peptide substrate binds across the lateral opening of the PflAE core, making contacts mainly to protein residues originating outside of the core, with some exceptions (Figure IV.17.a). The loop preceding the core plays a major part of substrate binding in PflAE, providing the critical interaction between $D_{15}$ and the amino group of the peptide glycine residue (which corresponds to Pfl $G_{734}$), as well as steric interactions with the peptide (Figure IV.19). This loop, which carries a motif specific to all of the AdoMet radical activases, (D/N)GxGxR, undergoes a major conformational change, swinging up into the active site and becoming fully ordered to properly bind substrate. The arginine of this motif interacts with neighboring $\beta$ strands, preserving the structural integrity of the AdoMet radical core during the loop's conformational change. Within the core, two interactions are provided by the $\beta$ strands – $F_{25}$, which provides steric interactions, and $R_{166}$, which is not located close enough to make specific contacts to the substrate, but may play a role in catalysis (Section IV.F.3.b). The only residues of the substrate binding site that are from within the core but are not located on a $\beta$ strand are $H_{37}$ and $N_{38}$, which are both found on the cluster-binding loop, directly after the CX3CX$\phi$C motif. Again, these residues provide specific interactions with the substrate at a critical place, the carbonyl group of the peptidic $G_{734}$, and appear key to maintaining the substrate conformation for hydrogen atom abstraction. Finally, residues $L_{204}$, $H_{207}$ and $K_{208}$ make up the remainder of the substrate binding site. They are all located on a loop immediately following the core that folds down over the cluster on the C-terminal side of the core's $\beta$ sheet. $L_{204}$ and $H_{207}$ are involved in recognition of a tyrosine residue

of the peptide substrate, while $K_{208}$ appears to help control the substrate conformation through an interaction to a peptide backbone atom (Figure IV.19). Overall, the coordination observed in the peptide-bound PflAE structure results in a distance of 4.1 Å between the AdoMet C5' atom and the Cα of $G_{734}$. The main aspects of substrate binding by PflAE identified by this structure are without a doubt the major interactions made in the full Pfl/PflAE complex, but further biochemical and structural studies of the full complex will provide a more detailed picture of these and any additional interactions.

### IV.G.2.b. HemN

The substrate of HemN is coproporphyrinogen III, a heme precursor (Section IV.C.2). However, the HemN structure was solved in the absence of the actual substrate; as discussed above, a second molecule of AdoMet (SAM2) is observed where the substrate would be expected to bind (Figures IV.16 and IV.18). Since there is no structure of HemN bound to its natural substrate, the binding site described here is that of SAM2. One side of the SAM2 binding site is made up by the first few strands of the core (β1-β4). Structural motifs from within the C-terminal region form the other side (Figure IV.17.b). Specifically, some important residues in the binding site (Figure IV.20) originate in the loop of the β-finger motif that follows the C-terminal HemN domain and a β strand that extends the HemN core β sheet at its N-terminus (orange in Figure IV.17.b, light pink ribbons and carbons in Figure IV.20). Because the actual substrate is not observed here and is much larger in size than AdoMet, the actual substrate binding site likely extends further along the AdoMet radical core.

Similar to the cluster-bound AdoMet, the adenine of SAM2 is recognized at the N6 and N7 atoms by interactions with both backbone atoms of $I_{329}$, which may also provide hydrophobic interactions (Figure IV.20). $Y_{56}$ of strand $\beta1$ and $F_{310}$ of the $\beta$ finger loop provide further hydrophobic interactions. The SAM2 ribose makes hydrogen bonds to $G_{112}$ and $E_{145}$ of the core (from the loop following $\beta2$ and within $\beta3$, respectively). No specific contacts to the methionine are observed, probably giving rise to the disorder observed in this part of the molecule[30]. Other notable residues in this area are $H_{58}$ and $Q_{311}$, which are essential to catalysis[36,47] but do not appear to make direct contacts to SAM2. The actual substrate likely makes more specific contacts, especially to these particular residues. $F_{310}$ and $Q_{311}$ are part of a HemN-specific motif, $_{308}KNFQGYTT_{315}$, which corresponds to the $\beta$-finger motif. The $\beta$-finger itself may therefore be an important substrate-binding motif for HemN, possibly playing a role in any conformational changes associated with substrate binding.

### IV.G.2.c. BioB

Biotin synthase catalyzes the insertion of a sulfur atom into dethiobiotin to form biotin (Section IV.C.3). The structure of BioB was solved in the presence of the actual substrate dethiobiotin, with the 4Fe-4S cluster in the oxidized state in order to prevent turnover[22-24]. In addition to the canonical 4Fe-4S cluster, a 2Fe-2S cluster is observed within the TIM barrel, serving as the source of sulfur for turnover[40,59]. As apparent from Figure IV.17.c, the substrate binding site is formed by residues from the middle of most of the barrel's $\beta$ strands, as is the 2Fe-2S cluster binding site. Several residues from loops following the $\beta$ strands make important interactions with the substrate as well.

The six-carbon "tail" of dethiobiotin is held through the carboxylate by BioB backbone atoms of residues $L_{290}$, $L_{291}$, $T_{292}$ and $T_{293}$, and by the hydroxyl of $T_{292}$ (Figure IV.21). The leucine sidechains, along with the sidechains of $L_{45}$ and $F_{285}$, provide a small hydrophobic surface for the carbon chain of dethiobiotin to pack against. Three conserved asparagines, $N_{151}$, $N_{153}$, and $N_{222}$, together ligate the ureido ring of dethiobiotin through hydrogen bonds to the nitrogens and carbonyl oxygen. $N_{151}$ and $N_{153}$ are part of a conserved $_{150}$YNHNLD$_{155}$ motif found in biotin synthases, which when mutated yield inactive enzyme[80]. These interactions hold dethiobiotin close to AdoMet, at a distance of 4.1 Å and 4.0 Å between C5' of AdoMet and C6 and C8 of the substrate, respectively. Finally, the 2Fe-2S cluster bound within the BioB barrel is coordinated by three cysteine residues ($C_{97}$, $C_{128}$ and $C_{188}$) and $R_{260}$ (Figure IV.21b), which may aid in reconstitution of the 2Fe-2S cluster or in modifying the characteristics of the cluster to aid in sulfur transfer.

### IV.G.2.d. MoaA

5'-GTP is the substrate of MoaA/MoaC, which together catalyze the complicated rearrangement of this substrate shown in Figure IV.1.b. The structure of MoaA in complex with 5'-GTP and the AdoMet cleavage products 5'-dA and methionine was solved by soaking 5'-GTP-free crystals in a solution containing 5'-GTP (Section IV.C.4). The second 4Fe-4S cluster of MoaA is also observed in this structure, ligating 5'-GTP through the nitrogens of the purine base. The 5'-GTP binding site is very extensive, and is formed by all of the β sheets of the AdoMet radical core (Figure IV.17.d). The C-terminal region of MoaA (colored grey in Figure IV.17.d) that extends the core β sheet on both sides is involved in substrate binding. Specific elements that form this side of

the binding site are several β strands outside of the core and some loops in the C-terminal region (highlighted orange in Figure IV.17.d).

The triphosphate of 5'-GTP is very extensively coordinated, mainly by active site arginines and lysines (Figure IV.22), including $R_{17}$, $K_{69}$, $R_{71}$, $K_{163}$, $T_{102}$, $N_{124}$, $N_{165}$ and $R_{192}$. Specific contacts to the guanine base are made by $R_{17}$, $R_{266}$ and $R_{268}$, while hydrophobic interactions are provided by $I_{194}$, $I_{253}$, $M_{197}$, $F_{260}$ and $L_{279}$. One interesting point that becomes apparent in an analysis of Figures IV.16 and IV.18 is that the substrate of MoaA is located furthest away from the C5' atom of AdoMet than any other substrate. However, four facts should be kept in mind while examining this structure: (1) AdoMet is cleaved in this structure, though substrate has not been transformed, possibly affecting the substrate binding site conformation, (2) MoaC, which is required for catalysis, is absent from the structure and (3) the enzyme was crystallized without 5'-GTP, which may prevent any necessary large conformational changes. Unknown conformational changes may occur upon binding of MoaC that could shift substrate closer to AdoMet. Finally, (4) because little is known about the specific mechanistic steps of this reaction, we can not yet be sure that 5'-GTP is the actual substrate of MoaA. Though it is likely that MoaA catalyzes the rearrangement reaction and MoaC is involved in pyrophosphate release[3,33], at this point we cannot rule out the possibility that the MoaA substrate is actually formed by MoaC.

The second 4Fe-4S cluster binds to MoaA at the N-terminus of an α helix within the C-terminal domain. It is coordinated by three cysteine residues, $C_{261}$, $C_{264}$ and $C_{278}$, which are all located on loops, again within the C-terminal domain of MoaA.

*IV.G.2.e. LAM*

The binding site provided by LAM for lysine and PLP is formed by several parts of the enzyme. First, the loop connecting the N-terminal helical domain and the AdoMet radical core provides part of the binding site (orange and grey in Figure IV.17.e). Some strands from the core are involved in binding substrate, as is the cluster-binding loop. Finally, protein elements outside of the core, such as a β finger motif and the loop following that motif provide the remainder of the binding site. Some residues from a second molecule of LAM provide interactions; specifically, the loop following the β finger motif of one molecule interacts with the PLP of a second molecule (Figure IV.17.e). The cofactor, PLP, is extensively coordinated.

The substrate lysine molecule is bound by four residues in the LAM active site: the carboxyl group is hydrogen bonded to $R_{134}$ and $S_{169}$, from the cluster binding loop and strand β2, respectively (Figure IV.23.a). $D_{293}$ and $D_{330}$ both ligate the amino group of the substrate, while $Y_{290}$ provides hydrophobic interactions with the four-carbon chain. These interactions together hold the substrate tightly in place, resulting in a distance of 3.8 Å between the AdoMet C5′ and Cβ of lysine.

In the absence of substrate, PLP forms an imine linkage to $K_{337}$[81], which is apparent in the structure[9] (Figure IV.23.b). Interestingly, the pyridoxal nitrogen interacts with only a water molecule ligated by the backbone atoms of $R_{112}$, $Y_{113}$ and $R_{116}$, a rare observation in PLP-dependent reactions. Typically the PLP pyridoxal nitrogen is hydrogen bonded by asparate, serine, arginine or glutamate[82-84]. Residues $E_{100}$ and $R_{112}$ bind the PLP via a water molecule, as does $Q_{258}$ of the GxIxGxxE motif. The phosphate group is coordinated by quite a few residues, including $R_{116}$, $R_{198}$, $Y_{287}$, $Y_{288}$, the water molecule ligated by $Q_{258}$, and $G_{321}$ of the second LAM molecule. Hydrophobic interactions with the pyridine ring are provided by $L_{118}$. Finally, the sidechain of $K_{337}$ is

held in place to reform the imine linkage by the hydroxyl group of $Y_{113}$. PLP is likely essential for proper coordination and orientation of substrate lysine, as without it, the other interactions provided by the protein to the lysine substrate would not be sufficient to overcome the conformational flexibility of this long chained amino acid.

*IV.G.3. Conformational changes associated with substrate binding in AdoMet radical enzymes*

As discussed above, the radical reactions catalyzed by these enzymes would likely require sequestration of the active sites from solvent during catalysis. This implies that conformational changes will occur upon substrate binding in most of the AdoMet radical enzymes, as has been observed for PflAE[51]. Specifically, the HemN and MoaA active sites appear very solvent exposed[30,33] (Figures IV.5 and IV.7). In the case of MoaA, no major conformational change is observed between the substrate-free and substrate-bound forms, leaving the active site very open[32,33]. However, an additional protein, MoaC, is required for activity, and its function during catalysis is as of yet unclear. It is possible that MoaC helps to bury the active site in some way, either directly or by enabling a conformational change. Obviously, further structural studies are required to identify conformational changes associated with substrate binding in the AdoMet radical enzymes, as only PflAE and MoaA have more than one structure available for identification of changes.

*IV.G.4. Similarities and differences in the substrate binding sites in AdoMet radical enzymes*

As has been emphasized in this section, each of these enzymes (with the exception of HemN) displays an ability to precisely control the conformation of its substrate (Figures IV.19 – IV.23). PflAE provides specific interactions to the peptide substrate and in particular to $G_{734}$, BioB and LAM use multiple hydrogen bonds to stretch their substrates across the active site and MoaA uses a second 4Fe-4S cluster and extensive hydrogen bonds to coordinate 5'-GTP. However, outside of this, there are virtually no similarities between the substrate binding sites. The enzymes have very different protein motifs outside of the AdoMet radical core, and the size and orientation of substrate binding sites are also dissimilar. There is no similarity noted in the locations of bound cofactors (Figure IV.16). This variation is necessary in order to ensure that the substrate binds properly to allow the correct hydrogen atom abstraction. From a structural point of view, the substrate conformation is certainly one of the most important aspects of catalysis. The highly divergent substrate binding sites discussed here reflect the diversity of the substrates and reactions of the AdoMet radical enzymes.

## IV.H. Reductant binding in AdoMet radical enzymes

The physiological reductants for most of these enzymes are flavodoxin and ferredoxin, and they serve to reduce the 4Fe-4S cluster from the 2+ to the 1+ state. No biochemical experiments aimed at determining where the reductant interacts with the AdoMet radical enzymes have been conducted; however, since so many other aspects of the initial radical generation are conserved, the general location of the reductant binding site should be similar in each enzyme. Our laboratory[51] and that of Dieter Jahn[30] have proposed the surface nearest the cluster-binding loop as this interface, and mapping the sequence conservation of each enzyme onto its surface supports this theory (Figure IV.24).

The surface proposed here for interaction with flavodoxin or another electron transfer protein is formed mainly by the cluster-binding loop, with additional surface added in some cases by the loops following strands β2 and β4. The loop following β2 seems to form part of the conserved surface of both PflAE (Figure IV.24.a) and BioB (Figure IV.24.c), while that following β4 (along with the N-terminal residues of helix α4A) forms part of the surface in HemN (Figure IV.24.b) and BioB. However, the pattern of high conservation on this particular surface is not present on the surface of LAM (Figure IV.24.e). The N-terminal helical domain of LAM (Section IV.D.2.e) sits atop the partial barrel above the 4Fe-4S cluster, obstructing the putative reductant-binding interface. It is not clear whether a conformational change could expose more of this surface, or a different surface on LAM interacts with the reductant.

The particular surface discussed here is an ideal candidate for forming the surface of interaction with the physiological reductant. The proximity of this surface to the cluster makes this location desirable –this surface is typically the closest region to the cluster (Section IV.E.2), within 7 to 10 Å in each of the structures, which is a reasonable distance for electron transfer[85]. Also, in order to prevent nonproductive cleavage of AdoMet, one would expect the reduction of the 4Fe-4S cluster to happen after substrate binds, when the active site is ready for catalysis. The location of this particular surface, on the opposite side of the enzyme from the lateral opening of the AdoMet radical core and active site, would allow retention of the reductant-enzyme interaction even with changes at the active site, from substrate binding during catalysis to evolution of the active site with preservation of the reductant binding site.

## IV.I. Conclusions

The currently available AdoMet radical structures provide an excellent cross-section of the superfamily and describe the protein fold used for radical generation in great detail. These structures have allowed an analysis of the similarities and differences between the enzymes in terms of the core fold, AdoMet binding sites, overall AdoMet conformation, and tailoring of the protein chain to specific substrates. Structure based sequence alignments can now be used to predict the residues involved in AdoMet binding and catalysis in other superfamily members. However, although several AdoMet radical structures are available, further structural study is required of this family in order to more clearly determine patterns of substrate binding and to clarify the structural questions that remain concerning the five enzymes studied here. Examples of such questions include the structure of the full PflAE/Pfl complex; how HemN binds the coproporphyrinogen III substrate and whether SAM2 is physiologically relevant; whether conformational changes occur in MoaA upon substrate binding; whether MoaC plays a structural role in this reaction or catalyzes a distinct reaction; whether any conformational changes are associated with substrate binding and release in BioB and LAM and where and how the physiological reductant interacts with each AdoMet radical enzyme. Clearly, there is much to be learned about this new and exciting class of enzymes.

**IV.J. References**

[1]  Sofia, H. J. et al., Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **29** (5), 1097 (2001).

[2]  Frey, P. A., Lysine 2,3-aminomutase: is adenosylmethionine a poor man's adenosylcobalamin? *FASEB J* **7** (8), 662 (1993).

3   Marquet, A., Tse Sum Bui, B., Smith, A. G., and Warren, M. J., Iron-sulfur proteins as initiators of radical chemistry. *Nat Prod Rep* **24** (5), 1027 (2007).

4   Wang, S. C. and Frey, P. A., S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem Sci* **32** (3), 101 (2007).

5   Dekker, E. E. and Barker, H. A., Identification and cobamide coenzyme-dependent formation of 3,5-diaminohexanoic acid, an intermediate in lysine fermentation. *J Biol Chem* **243** (12), 3232 (1968).

6   Morley, C. G. and Stadtman, T. C., Studies on the fermentation of D-alpha-lysine. Purification and properties of an adenosine triphosphate regulated B 12-coenzyme-dependent D-alpha-lysine mutase complex from Clostridium sticklandii. *Biochemistry* **9** (25), 4890 (1970).

7   Chirpich, T. P., Zappia, V., Costilow, R. N., and Barker, H. A., Lysine 2,3-aminomutase. Purification and properties of a pyridoxal phosphate and S-adenosylmethionine-activated enzyme. *J Biol Chem* **245** (7), 1778 (1970).

8   Frey, P. A., Ballinger, M. D., and Reed, G. H., S-adenosylmethionine: a 'poor man's coenzyme B12' in the reaction of lysine 2,3-aminomutase. *Biochem Soc Trans* **26** (3), 304 (1998).

9   Lepore, B. W., Ruzicka, F. J., Frey, P. A., and Ringe, D., The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale. *Proc Natl Acad Sci U S A* **102** (39), 13819 (2005).

10  Blakley, R. L., Ribonucleoside triphosphate reductase from Lactobacillus leichmannii. *Methods Enzymol* **51**, 246 (1978).

11  Tamarit, J. et al., The anaerobic ribonucleotide reductase from Escherichia coli. The small protein is an activating enzyme containing a [4fe-4s](2+) center. *J Biol Chem* **274** (44), 31291 (1999).

12  Magnusson, O. T., Reed, G. H., and Frey, P. A., Characterization of an allylic analogue of the 5'-deoxyadenosyl radical: an intermediate in the reaction of lysine 2,3-aminomutase. *Biochemistry* **40** (26), 7773 (2001).

13  Frey, M., Rothe, M., Wagner, A. F., and Knappe, J., Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. *J Biol Chem* **269** (17), 12432 (1994).

14  Wagner, A. F. et al., A dehydroalanyl residue can capture the 5'-deoxyadenosyl radical generated from S-adenosylmethionine by pyruvate formate-lyase-activating enzyme. *Biochem Biophys Res Commun* **254** (2), 306 (1999).

15  Fontecave, M., Mulliez, E., and Ollagnier-de-Choudens, S., Adenosylmethionine as a source of 5'-deoxyadenosyl radicals. *Curr Opin Chem Biol* **5** (5), 506 (2001).

16  Holliday, G. L. et al., Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep* **24** (5), 972 (2007).

17  Eschenmoser, A., Vitamin B12: Experiments Concerning the Origin of Its Molecular Structure. *Angewandte Chemie International Edition in English* **27** (1), 5 (1988).

18  Nicolet, Y. and Drennan, C. L., AdoMet radical proteins--from structure to evolution--alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* **32** (13), 4015 (2004).

[19] Kozbial, P. Z. and Mushegian, A. R., Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol* **5**, 19 (2005).

[20] Reichard, P., The evolution of ribonucleotide reduction. *Trends Biochem Sci* **22** (3), 81 (1997).

[21] Stubbe, J., Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr Opin Struct Biol* **10** (6), 731 (2000).

[22] Petrovich, R. M., Ruzicka, F. J., Reed, G. H., and Frey, P. A., Characterization of iron-sulfur clusters in lysine 2,3-aminomutase by electron paramagnetic resonance spectroscopy. *Biochemistry* **31** (44), 10774 (1992).

[23] Lieder, K. W. et al., S-Adenosylmethionine-dependent reduction of lysine 2,3-aminomutase and observation of the catalytically functional iron-sulfur centers by electron paramagnetic resonance. *Biochemistry* **37** (8), 2578 (1998).

[24] Broderick, J. B. et al., Pyruvate formate-lyase-activating enzyme: strictly anaerobic isolation yields active enzyme containing a [3Fe-4S](+) cluster. *Biochem Biophys Res Commun* **269** (2), 451 (2000).

[25] Krebs, C. et al., Coordination of adenosylmethionine to a unique iron site of the [4Fe-4S] of pyruvate formate-lyase activating enzyme: a Mossbauer spectroscopic study. *J Am Chem Soc* **124** (6), 912 (2002).

[26] Cosper, M. M. et al., The [4Fe-4S](2+) cluster in reconstituted biotin synthase binds S-adenosyl-L-methionine. *J Am Chem Soc* **124** (47), 14006 (2002).

[27] Cosper, M. M. et al., Structural studies of the interaction of S-adenosylmethionine with the [4Fe-4S] clusters in biotin synthase and pyruvate formate-lyase activating enzyme. *Protein Sci* **12** (7), 1573 (2003).

[28] Chen, D., Walsby, C., Hoffman, B. M., and Frey, P. A., Coordination and mechanism of reversible cleavage of S-adenosylmethionine by the [4Fe-4S] center in lysine 2,3-aminomutase. *J Am Chem Soc* **125** (39), 11788 (2003).

[29] Henshaw, T. F., Cheek, J., and Broderick, J. B., The [4Fe-4S]1+ Cluster of Pyruvate Formate-Lyase Activating Enzyme Generates the Glycyl Radical on Pyruvate Formate-Lyase: EPR-Detected Single Turnover. *J. Am. Chem. Soc.* **122** (34), 8331 (2000).

[30] Layer, G. et al., Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J* **22** (23), 6214 (2003).

[31] Berkovitch, F. et al., Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* **303** (5654), 76 (2004).

[32] Hanzelmann, P. and Schindelin, H., Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc Natl Acad Sci U S A* **101** (35), 12870 (2004).

[33] Hanzelmann, P. and Schindelin, H., Binding of 5'-GTP to the C-terminal FeS cluster of the radical S-adenosylmethionine enzyme MoaA provides insights into its mechanism. *Proc Natl Acad Sci U S A* **103** (18), 6829 (2006).

[34] Suzuki, Y. et al., Crystal Structure of the Radical SAM Enzyme Catalyzing Tricyclic Modified Base Formation in tRNA. *J Mol Biol* **372** (5), 1204 (2007).

[35] Goto-Ito, S. et al., Structure of an archaeal TYW1, the enzyme catalyzing the second step of wye-base biosynthesis. *Acta Crystallogr D Biol Crystallogr* **63** (Pt 10), 1059 (2007).

[36]  Layer, G., Verfurth, K., Mahlitz, E., and Jahn, D., Oxygen-independent coproporphyrinogen-III oxidase HemN from Escherichia coli. *J Biol Chem* **277** (37), 34136 (2002).

[37]  Layer, G. et al., Radical S-adenosylmethionine enzyme coproporphyrinogen III oxidase HemN: functional features of the [4Fe-4S] cluster and the two bound S-adenosyl-L-methionines. *J Biol Chem* **280** (32), 29038 (2005).

[38]  Layer, G. et al., The substrate radical of Escherichia coli oxygen-independent coproporphyrinogen III oxidase HemN. *J Biol Chem* **281** (23), 15727 (2006).

[39]  Tse Sum Bui, B. et al., Escherichia coli biotin synthase produces selenobiotin. Further evidence of the involvement of the [2Fe-2S]2+ cluster in the sulfur insertion step. *Biochemistry* **45** (11), 3824 (2006).

[40]  Ugulava, N. B., Surerus, K. K., and Jarrett, J. T., Evidence from Mossbauer spectroscopy for distinct [2Fe-2S](2+) and [4Fe-4S](2+) cluster binding sites in biotin synthase from Escherichia coli. *J Am Chem Soc* **124** (31), 9050 (2002).

[41]  Choi-Rhee, E. and Cronan, J. E., Biotin synthase is catalytic in vivo, but catalysis engenders destruction of the protein. *Chem Biol* **12** (4), 461 (2005).

[42]  Wuebbens, M. M. and Rajagopalan, K. V., Investigation of the early steps of molybdopterin biosynthesis in Escherichia coli through the use of in vivo labeling studies. *J Biol Chem* **270** (3), 1082 (1995).

[43]  Rieder, C. et al., Rearrangement reactions in the biosynthesis of molybdopterin-- an NMR study with multiply 13C/15N labelled precursors. *Eur J Biochem* **255** (1), 24 (1998).

[44]  Wuebbens, M. M., Liu, M. T., Rajagopalan, K., and Schindelin, H., Insights into molybdenum cofactor deficiency provided by the crystal structure of the molybdenum cofactor biosynthesis protein MoaC. *Structure* **8** (7), 709 (2000).

[45]  Wang, S. C. and Frey, P. A., Binding Energy in the One-Electron Reductive Cleavage of S-Adenosylmethionine in Lysine 2,3-Aminomutase, a Radical SAM Enzyme. *Biochemistry* (2007).

[46]  Layer, G., Heinz, D. W., Jahn, D., and Schubert, W. D., Structure and function of radical SAM enzymes. *Curr Opin Chem Biol* **8** (5), 468 (2004).

[47]  Layer, G. et al., Structural and functional comparison of HemN to other radical SAM enzymes. *Biol Chem* **386** (10), 971 (2005).

[48]  Nicolet, Y. et al., Desulfovibrio desulfuricans iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* **7** (1), 13 (1999).

[49]  Sanyal, I., Cohen, G., and Flint, D. H., Biotin synthase: purification, characterization as a [2Fe-2S]cluster protein, and in vitro activity of the Escherichia coli bioB gene product. *Biochemistry* **33** (12), 3625 (1994).

[50]  Markham, G. D., Norrby, P. O., and Bock, C. W., S-adenosylmethionine conformations in solution and in protein complexes: conformational influences of the sulfonium group. *Biochemistry* **41** (24), 7636 (2002).

[51]  Vey, J. Yang, J., Li, M., Broderick, W., Broderick, J., Drennan, C.

[52]  Cosper, N. J. et al., Direct FeS cluster involvement in generation of a radical in lysine 2,3-aminomutase. *Biochemistry* **39** (51), 15668 (2000).

[53]  Jarrett, J. T., The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. *Curr Opin Chem Biol* **7** (2), 174 (2003).

54 Seehra, J. S., Jordan, P. M., and Akhtar, M., Anaerobic and aerobic coproporphyrinogen III oxidases of Rhodopseudomonas spheroides. Mechanism and stereochemistry of vinyl group formation. *Biochem J* **209** (3), 709 (1983).

55 Ollagnier-de-Choudens, S., Mulliez, E., and Fontecave, M., The PLP-dependent biotin synthase from Escherichia coli: mechanistic studies. *FEBS Lett* **532** (3), 465 (2002).

56 Ollagnier-de Choudens, S. et al., Reductive cleavage of S-adenosylmethionine by biotin synthase from Escherichia coli. *J Biol Chem* **277** (16), 13449 (2002).

57 Guianvarc'h, D. et al., Biotin synthase, a new member of the family of enzymes which uses S-adenosylmethionine as a source of deoxyadenosyl radical. *Biochem Biophys Res Commun* **236** (2), 402 (1997).

58 Escalettes, F. et al., Biotin Synthase Mechanism: Evidence for Hydrogen Transfer from the Substrate into Deoxyadenosine. *J. Am. Chem. Soc.* **121** (15), 3571 (1999).

59 Lotierzo, M. et al., Biotin synthase mechanism: an overview. *Biochem Soc Trans* **33** (Pt 4), 820 (2005).

60 Cicchillo, R. M. et al., Lipoyl synthase requires two equivalents of S-adenosyl-L-methionine to synthesize one equivalent of lipoic acid. *Biochemistry* **43** (21), 6378 (2004).

61 Cicchillo, R. M. and Booker, S. J., Mechanistic investigations of lipoic acid biosynthesis in Escherichia coli: both sulfur atoms in lipoic acid are contributed by the same lipoyl synthase polypeptide. *J Am Chem Soc* **127** (9), 2860 (2005).

62 Florentin, D. et al., On the mechanism of biotin synthase of Bacillus sphaericus. *C R Acad Sci III* **317** (6), 485 (1994).

63 Tse Sum Bui, B. et al., Further investigation on the turnover of Escherichia coli biotin synthase with dethiobiotin and 9-mercaptodethiobiotin as substrates. *Biochemistry* **43** (51), 16432 (2004).

64 Daley, C. J. and Holm, R. H., Reactivity of [Fe4S4(SR)4]2-,3- clusters with sulfonium cations: analogue reaction systems for the initial step in biotin synthase catalysis. *Inorg Chem* **40** (12), 2785 (2001).

65 Daley, C. J. and Holm, R. H., Reactions of site-differentiated [Fe4S4]2+, 1+ clusters with sulfonium cations: reactivity analogues of biotin synthase and other members of the S-adenosylmethionine enzyme family. *J Inorg Biochem* **97** (3), 287 (2003).

66 Vidgren, J., Svensson, L. A., and Liljas, A., Crystal structure of catechol O-methyltransferase. *Nature* **368** (6469), 354 (1994).

67 Cheng, X. et al., Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74** (2), 299 (1993).

68 Korolev, S. et al., The crystal structure of spermidine synthase with a multisubstrate adduct inhibitor. *Nat Struct Biol* **9** (1), 27 (2002).

69 Dong, C. et al., Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **427** (6974), 561 (2004).

70 Huang, C. C. et al., Crystal structures of mycolic acid cyclopropane synthases from Mycobacterium tuberculosis. *J Biol Chem* **277** (13), 11559 (2002).

71 Anantharaman, V., Koonin, E. V., and Aravind, L., SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies,

and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol* **4** (1), 71 (2002).

[72] Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M., Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* **8**, 73 (2007).

[73] Jenuwein, T., Laible, G., Dorn, R., and Reuter, G., SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* **54** (1), 80 (1998).

[74] Alvarez-Venegas, R. and Avramova, Z., SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene* **285** (1-2), 25 (2002).

[75] Demirci, H., Gregory, S. T., Dahlberg, A. E., and Jogl, G., Recognition of ribosomal protein L11 by the protein trimethyltransferase PrmA. *EMBO J* **26** (2), 567 (2007).

[76] Pilka, E.S., Shafqat, N., Kavanagh, K.L., Cooper, C., Hozjan, V., Turnbull, A., Von Delft, F., Arrowsmith, C.H., Edwards, A., Weigelt, J., Sundstrom, M., Oppermann, U., Structural Genomics Consortium (SGC) Crystal structure of the human S-adenosylmethionine synthetase 1 in complex with the product. . *To be published*.

[77] Capitani, G. et al., Structure of ACC synthase inactivated by the mechanism-based inhibitor L-vinylglycine. *FEBS Lett* **579** (11), 2458 (2005).

[78] Rafferty, J. B., Somers, W. S., Saint-Girons, I., and Phillips, S. E., Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor. *Nature* **341** (6244), 705 (1989).

[79] Dixon, M. M., Huang, S., Matthews, R. G., and Ludwig, M., The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* **4** (11), 1263 (1996).

[80] Lotierzo, M. et al., Biotin synthase mechanism: mutagenesis of the YNHNLD conserved motif. *Biochemistry* **45** (40), 12274 (2006).

[81] Chen, D. and Frey, P. A., Identification of lysine 346 as a functionally important residue for pyridoxal 5'-phosphate binding and catalysis in lysine 2, 3-aminomutase from Bacillus subtilis. *Biochemistry* **40** (2), 596 (2001).

[82] Jansonius, J. N., Structure, evolution and action of vitamin B6-dependent enzymes. *Curr Opin Struct Biol* **8** (6), 759 (1998).

[83] Denessiouk, K. A. et al., Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes. *Proteins* **35** (2), 250 (1999).

[84] Schneider, G., Kack, H., and Lindqvist, Y., The manifold of vitamin B6 dependent enzymes. *Structure* **8** (1), R1 (2000).

[85] Page, C. C., Moser, C. C., and Dutton, P. L., Mechanism for electron transfer within and between proteins. *Curr Opin Chem Biol* **7** (5), 551 (2003).

## IV.K. Tables and Figures

**Table IV.1: Available AdoMet radical structures.** [a]The structure HemN contains a mixture of AdoMet enantiomers. The physiologically relevant enantiomer was used to obtain the values in this table and to prepare all figures. [b]GTP-MoaA has the AdoMet cleavage products bound. In this structure, the 5'-dA amino group does not ligate the unique iron. A third MoaA structure, of the R17/266/268A triple mutant, has been omitted from this table. [c]TYW1 is a very recently solved AdoMet radical enzyme (coordinates not available at this time). This structure does not contain the 4Fe-4S cluster, AdoMet or substrate.

| Enzyme | PflAE | HemN[a] | BioB | MoaA | GTP-MoaA[b] | KAM | TYW1[c] | Spectro-scopic studies of PflAE |
|---|---|---|---|---|---|---|---|---|
| PDB ID | | 1OLT | 1R30 | 1TV8 | 2FB3 | 2A5H | 2Z2U | ——— |
| Resolution | 2.77 | 2.07 | 3.4 | 2.2 | 2.35 | 2.1 | 2.4 | ——— |
| R / Rfree (%) | 22.9 / 26.1 | 15.5 / 18.7 | 25.6 / 30.0 | 21.4 / 24.1 | 22.0 / 27.1 | 18.7 / 22.5 | 22.0 / 29.6 | ——— |
| Space group | P6122 | P63 | P3121 | P212121 | P212121 | C2 | unknown | ——— |
| Release date | | Dec-03 | Jan-04 | Aug-04 | | Oct-05 | HPUB | ——— |
| Source | E. coli | E. coli | E. coli | S. aureus | S. aureus | C. subtermi-nale SB4 | M. jannas-chii | ——— |
| Method | Fe MAD | Fe MAD | Fe MAD | MR with apo (Fe MAD) | MR with 1TV8 | Se MAD | Se MAD | ——— |
| Ligands | 4Fe-4S, AdoMet, peptide | 4Fe-4S, AdoMet | 4Fe-4S, AdoMet, DTB, 2Fe-2S | Two 4Fe-4S, AdoMet | Two 4Fe-4S, 5'dA, Met, GTP | 4Fe-4S, AdoMet, PLP | none | ——— |
| # chains / ASU | 1 | 1 | 2 | 2 | 2 | 4 | Unknown | ——— |
| Distances | | | | | | | | |
| Fe - AdoMet N | 2.2 | 2.6 | 2.4 | 2.3 | 2.8 | 2.0 | ——— | ——— |
| Fe - AdoMet O | 2.2 | 2.2 | 2.5 | 2.0 | 1.8 | 2.0 | ——— | ——— |
| Fe - AdoMet C | 3.0 | 3.1 | 3.0 | 2.9 | 2.7 | 3.0 | ——— | 3.3 |
| Fe - AdoMet S | 3.2 | 3.5 | 4.0 | 3.2 | 3.6 | 3.1 | ——— | ——— |
| Fe - AdoMet CH3 | 3.5 | 4.4 | 5.4 | 4.4 | 4.3 | 4.5 | ——— | 4-5 |
| S - AdoMet S | 3.9 | 3.6 | 4.2 | 3.4 | 4.1 | 3.9 | ——— | ——— |

**Figure IV.1: AdoMet radical reactions.** (a) The general reaction each AdoMet radical enzyme catalyzes to initiate radical chemistry. (b) Reactions catalyzed by the AdoMet radical enzymes for which structures are currently known: PflAE, HemN ($R_1$ and $R_2$ in this reaction scheme correspond to the remainder of the coproporphyrinogen III tetrapyrrole macrocycle), BioB, MoaA/MoaC, and LAM. Adapted from reference 4.
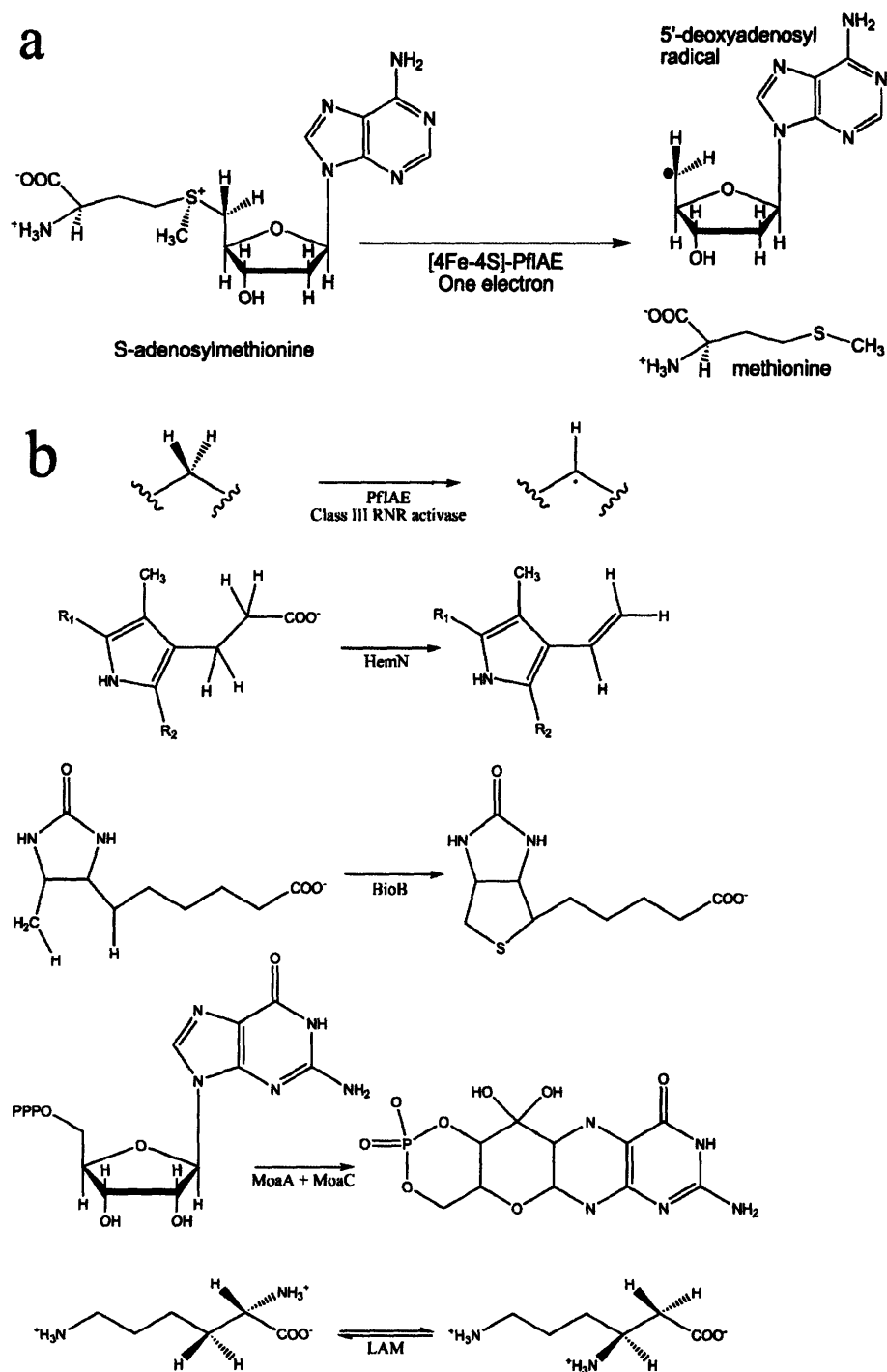
**Figure IV.2: Diagrams of the complete substrate of each structurally characterized AdoMet radical enzyme.** (a) PflAE substrate: Pfl (b) HemN substrate: coproporphyrinogen III (c) BioB substrate: dethiobiotin (d) MoaA's putative substrate: 5′-GTP (e) LAM substrate: lysine.
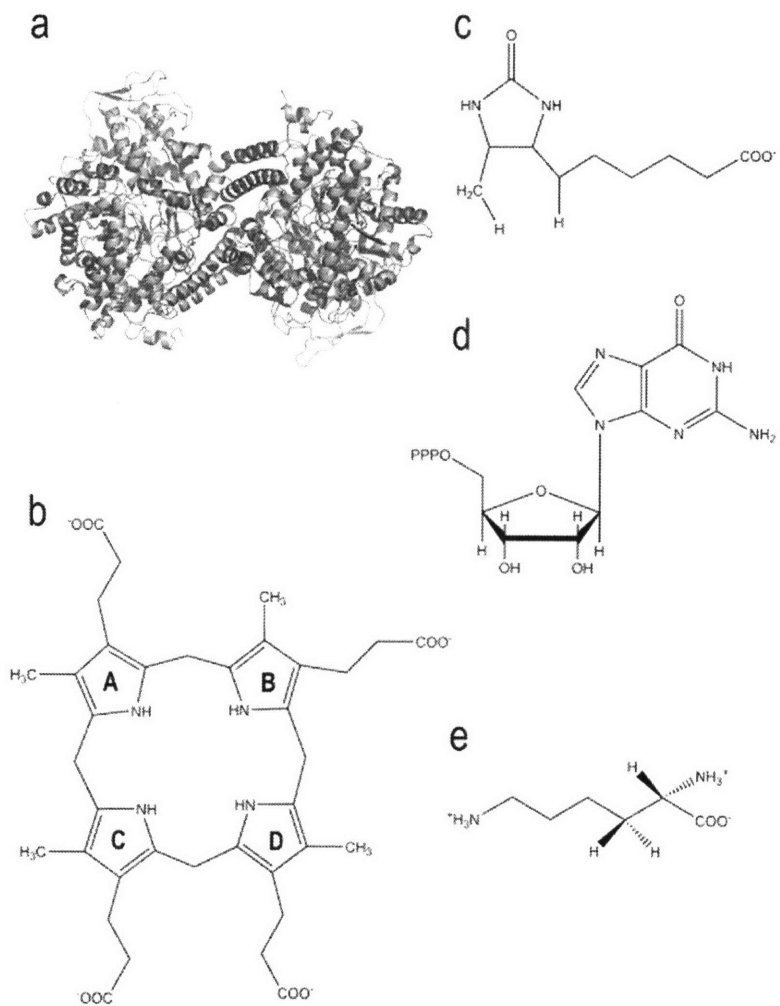
**Figure IV.3:  The (a) top and (b) side views of a typical TIM barrel fold**.  The enzyme shown is triosephosphate isomerase solved by X-ray crystallography to 0.83 Å resolution, PDB code 1N55.  β strands are colored yellow, and α helices are teal.
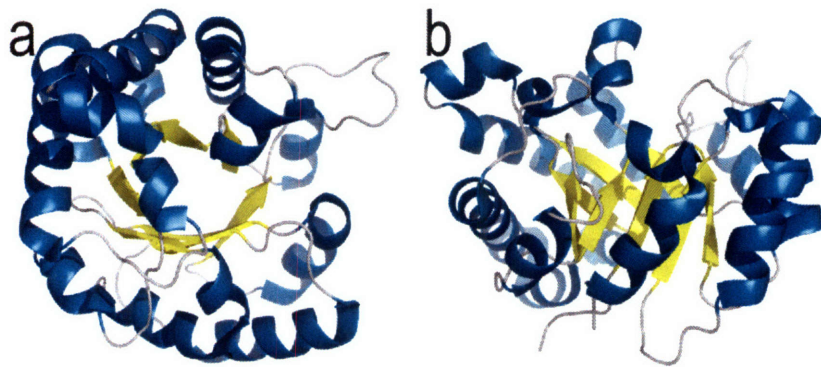
**Figure IV.4: Ribbon diagrams of PflAE.** Shown are the (a) top view and (b) side view of the monomer with the 4Fe-4S cluster, AdoMet and peptide substrate in stick representation. The AdoMet radical core domain is colored as follows: helices, teal; strands, yellow; loops, grey; cluster-binding loop harboring the $CX_3CX\phi C$ motif, magenta. The AdoMet, 4Fe-4S cluster and substrate atoms are colored as follows: iron, ruby; sulfur, gold; AdoMet carbons, green; substrate carbons, teal; oxygen, red, nitrogen, blue. Protein elements outside the core are colored grey. (c) The PflAE monomer shown in stereoview, with cartoons colored as a rainbow (blue at N-terminus, red at C-terminus). The cluster, AdoMet and substrate are colored as in (a) and shown in thick ball and sticks for clarity.
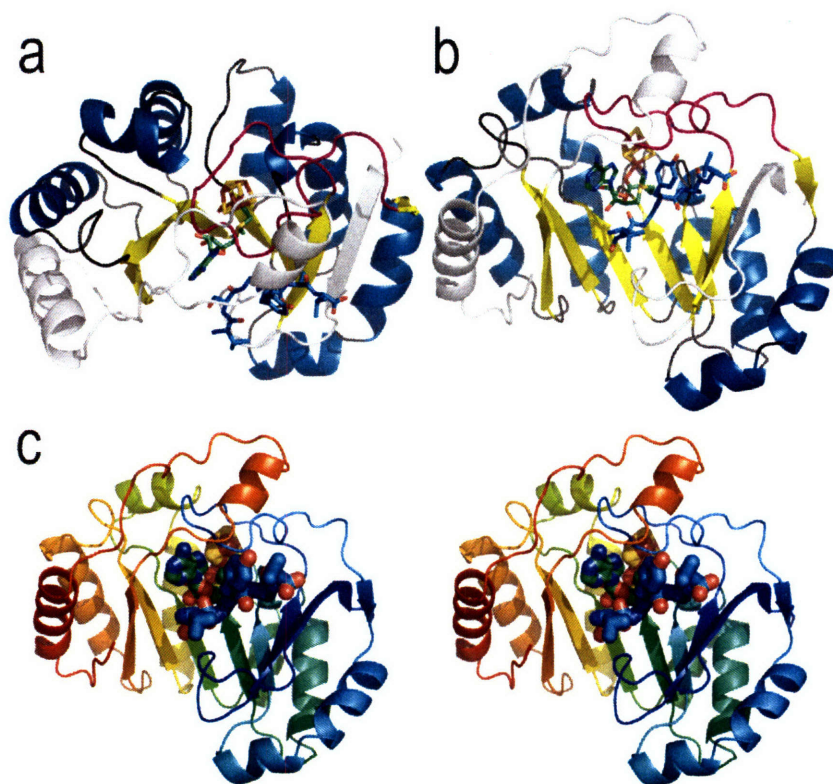
**Figure IV.5: Ribbon diagrams of HemN**. Shown are the (a) top view and (b) side view of the monomeric protein. The AdoMet radical core domain is colored as in Figure IV.4.a with the 4Fe-4S cluster and two AdoMet molecules shown in stick representation. (c) The HemN monomer shown in stereoview, displayed and colored as in Figure IV.4.c. Note that the N-terminus, colored blue, is partially disordered as it wraps around the molecule between the two domains (chain breaks and missing protein sequence).
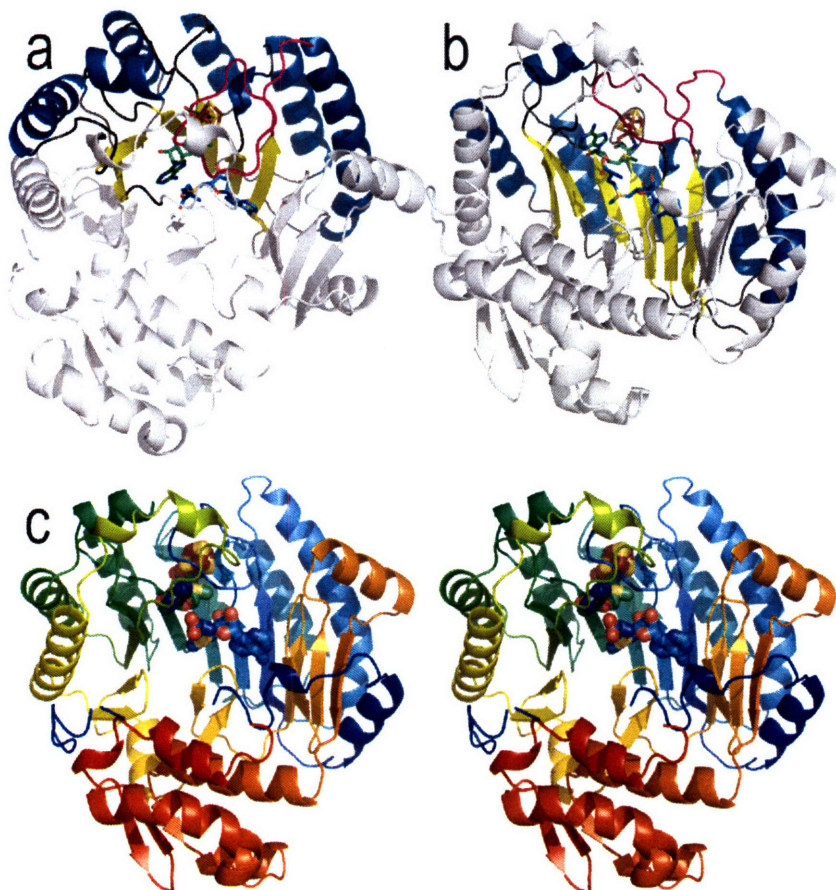
**Figure IV.6:  Ribbon diagrams of BioB**.  Shown are the (a) top view and (b) side view of the monomer.  The AdoMet radical core domain is colored as in Figure IV.4.a with the 4Fe-4S cluster, 2Fe-2S cluster, AdoMet and dethiobiotin shown in stick representation. Here, the elements outside of the core (shown in grey) complete the $(\beta/\alpha)_8$ barrel.  (c) The BioB dimer shown in stereoview, displayed and colored as in Figure IV.4.c.  The second molecule of the dimer is shown in light grey cartoons.
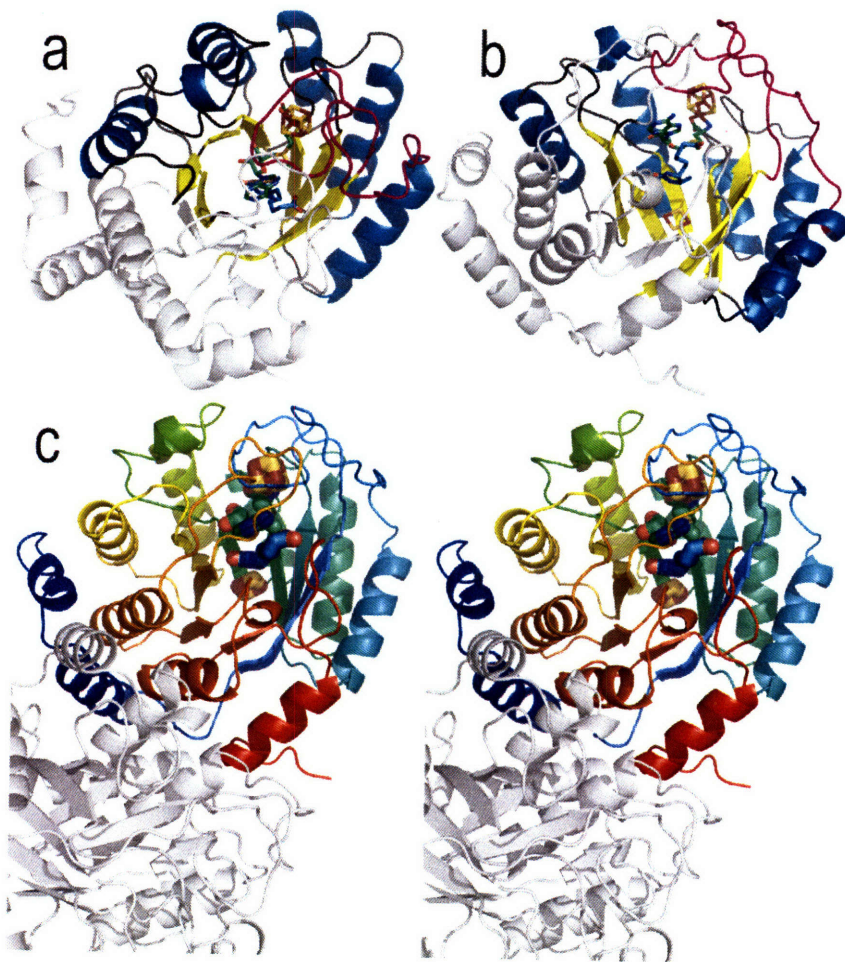
**Figure IV.7: Ribbon diagrams of MoaA.** Shown are the (a) top view and (b) side view of the monomer. The AdoMet radical core domain is displayed and colored as in Figure IV.4.a with the two 4Fe-4S clusters, AdoMet and 5′-GTP shown in stick representation. (c) The MoaA dimer shown in stereoview colored as in Figure IV.4.c.
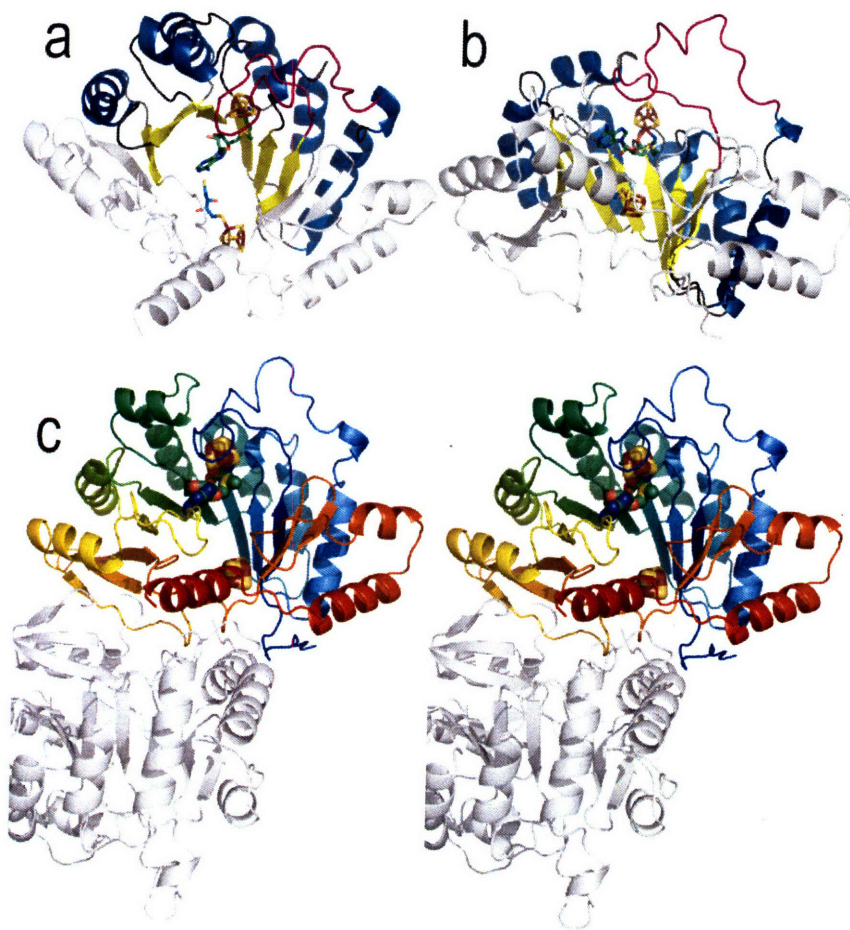
**Figure IV.8: Ribbon diagrams of LAM**. Shown are the (a) top view and (b) side view of the monomer. In (a) and (b), the AdoMet radical core domain is colored as in Figure IV.4.a., and part of a second LAM monomer is also shown in ribbons, colored peach. The 4Fe-4S cluster, AdoMet, PLP and lysine substrate are shown in stick representation. (c) The LAM oligomer shown in stereoview, displayed and colored as in Figure IV.4.c. The second molecule is shown in dark grey cartoons, while the two other molecules of the tetramer are shown in light grey cartoons. In order to help visualize the oligomerization surfaces, the tetramerization surface is marked "A" in parts a and c, and the dimerization surface "B" in parts b and c.
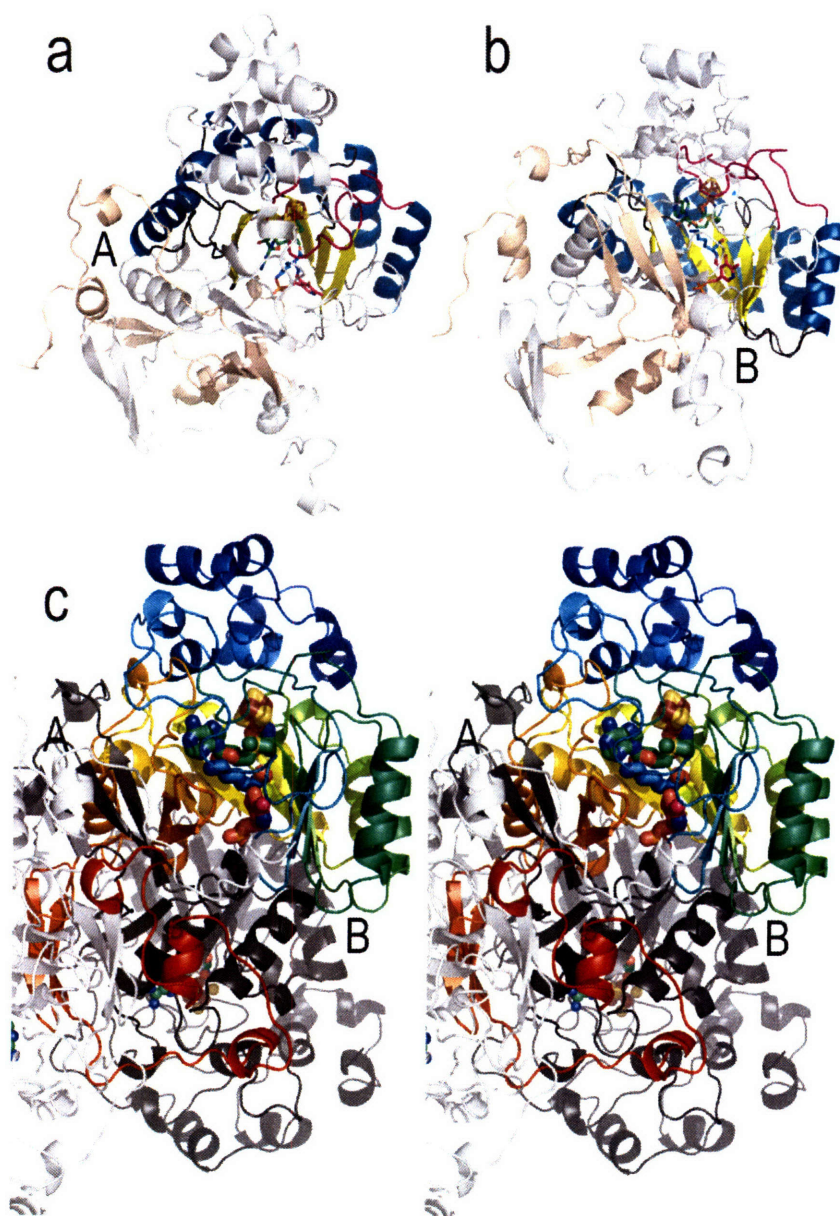
**Figure IV.9: Superposition of the five AdoMet radical core domains**, shown in stereo and colored as in panels *a* and *b* of Figures IV.4 – IV.8. Here are shown two side views of the AdoMet radical core, from (a) the front and (b) the back.
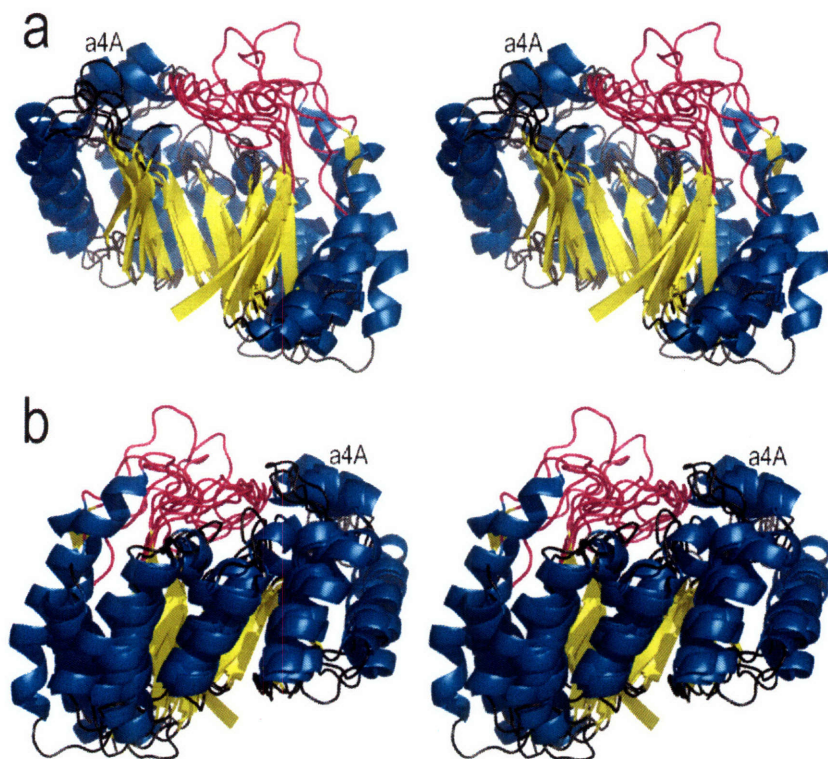
**Figure IV.10: Structure based sequence alignment of the AdoMet radical enzymes** (figure is on the following page). The five enzymes for which a structure is available are listed at the top of the alignment with the main secondary structural elements labeled at the top of the alignment, highlighted in yellow (strands) and teal (helices). Residues of interest are colored as follows: the $CX_3CX\phi C$ motif cysteines are in red; residues that contact AdoMet in green; residues that contact the substrate, blue; residues that contact a cofactor, pink; and residues that contact both AdoMet and the substrate, orange. The sequences of the first five enzymes are shown with no sequence gaps; however, the remaining sequences do have some gaps. Residues of the four motifs discussed in the text are boxed in red and identified as follows: red stars correspond to the cysteines of the cluster-binding loop; red circles, the GGE motif; red triangles, the positions within the GxIxGxxE motif; and red squares, the conserved structural motif. A previously published alignment[18] was used as a starting point for this alignment. The alignment was then adjusted manually to reflect the exact structural elements of each enzyme and, in the case of the GxIxGxxE motif, to align the residues involved in conserved hydrogen bonding networks.

Multiple sequence alignment figure.

Block 1 — secondary structure: β1, α1 (positions 1, 10, 20, 30, 40, 50, 60)

```
PflA_E.coli   .SVIGRIHSFESCGTVDGPGIRFITFF..QGCLMR..CLYCHNRDTWDTHGGKEVTVEDLMKEVVTYRHF
HemN_E.coli   SEDFGEQAFLQAVARYPERPLSLYVHI.PFCHKL..CYFCGCNSRYKTGLEAERLME...VEQVLESAR
BioB_E.coli   LLFEAQQVERQHFDPRQVCVSTLLSIK.TGACPED..CKYCFQSSRYKTGLEAERLME...VEQVLESAR
MoaA_E.coli   ..MVEQIKDKLGRPIRDLRLSVTDRCNFR..CDYCMPKEVGDDFVFLPKNELLTFDEMARIAK
KamA_C.subt   DPLHEDIDSPVPGLTRYPDRVLLLIT.DMCSMY..CRHCTRRRFA...GCSDDSMP..MERIDKAID
Unk1_M.jann   ......KKIEITSTIHI.TNICHVNPXCLYCGF.......GYYEFFRLTDEEIKKSAIAIE
ThiH_E.coli   ......SNLCAND..CTYCGF............LDEADIARESAAIR
Unk2_D.vulg   .....GEHVEVRGLIEF.SSHCVRD..CRYCGL............REVIDAARLAV
PylB_M.maze   .....GNRVFLNCFIYF.STYCKNQ..CSFCYY............LTGEEVKEMCKALK
CofG_M.jann   .....SXWCRNK..CGYCIF............NEVKEILLKGD
CofH_M.jann   .....TNICVGN..CRFCAF............FLDIDEIAKRAVEAK
LipA_E.coli   .....GTATFMILGAICTRR..CPHCDV............
MiaB_E.coli   .....AFVSI.MEGCNKY..CTYCVVP............
NifB_A.vine   AHHYFARMHVAV.APACNIQ..CHYCNR...GVVSEVLTPEQAVKKVKAVAA
BssD_T.arom   .DGPGIRTTIFL..KGCPLR..CPWCEN...VDEILREALSDS
NrdG_E.coli   .NGPGTRCTLFV..SGCVHE..CPGCYN............
NirJ_P.aeru   .....RCNLT..CKHCYS...AEILRGIDDLR
SplB_B.subt   .....GCMGH..CHYCYL..P.YIRTYVNVEEILDQADKYM
```

Block 2 — secondary structure: β2, α2, β3, α3 (positions 70, 80, 90, 100, 110)

```
PflA_E.coli   MKAS....GGGVTASG.GE...AILQAEFVRDWFRACK.KEGI...HTCLDTK.....GFVRRYDPVIDE
HemN_E.coli   AFLTAGRIVSQLLWG.GTPTYLNKAQISRLMKLLRINFQTNADAEISIEVD...PREIELDVLDDHL
BioB_E.coli   KAKAAG..STRFCMGA.AWKKPEERDMPYLEQMVQGVKAMG...LEACMTL...GT.LSSSQAQRL
MoaA_E.coli   VYAILG..VKKIRITG.GEPLMR...RDLDVLIAXLNQIDGIED IGLTTN...GLLEKKNGQRLY
KamA_C.subt   YLRNTPQ.VRDVLLSG.GDALLVS..DETREYIIAAKLREIPEVE..IVRLSRTPVVLPQRITPELVNML
Unk1_M.jann   ES...G..IKRVSCSS.AHGYQG...KEVIRALKIVKKYTN.LEVLVNA.....GADLIEESIKEL
ThiH_E.coli   EM...G..FEHLLVT.GEHQAKVGM..DYFRRHLPALREQFSSLQMEVQ....P..LAETEYAEL
Unk2_D.vulg   AG...G..ADTIVLQS.GDDLACR...TAAMTRGTRPELG.VVVTLGC.....GERPP.RDYALW
PylB_M.maze   GA...G..FHMIDLTM.GEDFY...RFVELVRTVKEELG.IPIMISP.....GV.MDDSILLKA
CofG_M.jann   RL...G..CREALFTF.GEHVDE........S.LLPUTNC.....GI.LNYDEL.KM
CofH_M.jann   KF...G..CTEVCIQG.GLHPK...EILKAVHEATKFYGDIHIHAFS.....LDIKEALKI
LipA_E.coli   LTAT.P..LRYVVITS.VDRDDL...HFADCITAIRPKSPQIKIETLV.....PD...DRALDT
MiaB_E.coli   .....G..VREVNDLG.........FADLLRLVAAIDGIDIRIRFTI.....FIDDIIEVY
NifB_A.vine   AI...PQ..MSVLGIAGPDPLA...RTLDTFRMLSEQAPDKKLCVSTN.....GLALP.ECVEEL
BssD_T.arom   AFYRN..SGGGVTIS.G.GDPLYF...PDFTRQLASELHARG.VHVAIETSC.....GFNDSHA..
NrdG_E.coli   .....QGISLSG.G.GDPLHFQNV...PDILKLVQRIRAECPGKDIWVWT.....GYKLDE......
NirJ_P.aeru   AA...G..VRVLILSG.GEPLMH...PDLFEIAABARQA.G.KFVALSSN.....GTLIDEGNIQRV
SplB_B.subt   KERA.P..EFTRFEASCTSDIVG...IDHLTHTLKRAIEHFGQSDLGKLRF.....VTKFHHVDHLLD
```

Block 3 — secondary structure: β4, α4a, α4, β5 (positions 120, 130, 140, 150, 160, 170, 180)

```
PflA_E.coli   ILE.....VTDIVMLDLKCMNDEIHQNLV.GVSNHRTLEFAKYLANKNVK...VWIRYVVVFGWSDDSA
HemN_E.coli   RAE..GFNRLSMGVQDFN..KEVQRLVNREDEEFSFALSNEAREI.GFTST..NIDLYGLPKQTPESFA
BioB_E.coli   AKA..GLD.YNHNLDFS..PEFYGNIITTRTYQERLDTLEKVRDA.CI..KVCSGCIVGLG.ETVKDRA
MoaA_E.coli   DA...GL.RRINVSLDAIDD.TLFQSINRNIKATTLEQIDYATSI.GL..NCVKVNVVIQKGINDDQIIP
KamA_C.subt   KKYH.P...VWLNTHFNH...PNEITEARACQLLAA.GVP..LGNCSVLLRGVNDCVHVX
Unk1_M.jann   KKY.G..IDTICCNLETINENLFKKVKF..GEELEDRIRVCKLVNKY.D..IELSTGLLIGI.GESYEDRV
ThiH_E.coli   KQL..G.LDGVKVVYQETYHE.......QDFFWRLETPDRLGRA.G..KIGLGALIGL.SDNWR..
Unk2_D.vulg   REA..G.ASRYLMKHETADAALYARLHF.GRTLEQRLEALRTLHGL.G..YEIGSSFIIGIPGQMPEVLA
PylB_M.maze   REE..G.ANFFALYQETYDRELYGKIRV.GSQSFEGRFNARRFAKEQ.G..LVKVNSVMIP.GVNDEHL..
CofG_M.jann   LK.D.V.NASMGLMLENASERLM..SF.GKHPKLRIEMIENAGKLK...IPFTTGLLIGI.GETNEEIV
CofH_M.jann   LKEN.G..........KEWIYIIKEAHKL.G..IPTTATMMYCHIEE.YKHWV
LipA_E.coli   EVMRDLRR...HGVTMLTLGQYLQPSR............
MiaB_E.coli   RDT.PELVSFLHLPVQSGSDRILNLMGRHTHALEYKAIIRKLRAAR.PD.IQISSDFIV.GFPGETTED
NifB_A.vine   AKHN...IDHVTIIINCVDPEI.......IEQQQKGEMLVAR.GILVKVNSVMIP.GVNDEHL..
BssD_T.arom   ....GIVDLFIVDLKTLDAHKELDVI..GWPLADILANLETLFAA.GAKVRIHDVIP.GFNDSHA..
NrdG_E.coli   ..QVVDLINVLVDG...........KEVQDLKDPSLIWR..G.......
NirJ_P.aeru   ALA..R.FDYVGISLDGL........SFDAALAAMRLCRFA.D.......
SplB_B.subt   AKHN.G..KTRFRFSINADYV......FLDKRIEAAVKVAKA.GYPLGF......
```

Block 4 — secondary structure: α5, β6 (positions 190, 200, 210, 220, 230, 240)

```
PflA_E.coli   HRLGEFTRDMGNV...EKIELLF.YHELGKHKWVAMGEEYKLDGVKPPKKETMERVKGILEQYGHKVMF.
HemN_E.coli   FTLKRVAELNPD.....RLSVFN.YAHLPTIFAAQRXIKKADLPSPQQKLDILQETIAFLTQSGYQFIGM
BioB_E.coli   GLLLQLANLFIPDE...SVPINM.LVKVKGTLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMN
MoaA_E.coli   MLEYFKLKEIE...IRFIE.FMDVGNDNGWDFSKVVTKDEKLIMIEQHFEIDPVEPKYFGEVAKYY
KamA_C.subt   KELVNKVKIRVRP....YYIYQ.CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGK
Unk1_M.jann   EHLFYLKNELNVGEIPIMGFNPYKGT.FMENHPKCSALEQAKTIAITRLIFPNIRIITSPTP..
ThiH_E.coli   AEHLLWLQQHY.....SVSFPRLRPCTG.GIEPASIMDERQLVQTICAFRLLAPEIELSIST.RE..
Unk2_D.vulg   DDILLAR...SLGVEMCGAGPFIPQAQTPLGGEPQGSTALALRVLAALRIALPDANLPATT..
PylB_M.maze   .....NPDMVRVMTFLPQEGTPLEGFSDSSKLSELKIIAILRLMFPECLI..
CofG_M.jann   DSLFKIKEI....IQEVIIQKFRAKKGIPMENFKEPSPIKMLKVIILAKLILDDISIQI..
CofH_M.jann   NULFIIKEIEI.......EDLKVFAVSRIIFKG..
LipA_E.coli   EVMRDLRR...HGVTMLTLGQYLQPSR..
MiaB_E.coli   FEKTMKLTADVNFDMSVSFITFSAR.PGTPA..
NifB_A.vine   ...........
BssD_T.arom   .....SGIDLLNFHCY..
NrdG_E.coli   ...........
NirJ_P.aeru   ...........
SplB_B.subt   ...........
```

Block 5 (positions continue)

```
PflA_E.coli   ...........
HemN_E.coli   DHFARPDDELAVAQREGVLHRNFQGYTTIQGDTLLLGMGVSAISM.............
BioB_E.coli   EQTQAMCFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNFQ.............
MoaA_E.coli   RHKDNGVQFGLITSVSQSFCSTCTRARLSSDGKFYGCLFATVDG.............
KamA_C.subt   TFVMPNYVISQSHDKVILRNFEGVITTYSEPINYTPGCNCDVCT.............
Unk1_M.jann   ...........
ThiH_E.coli   ...........
Unk2_D.vulg   ...........
PylB_M.maze   ...........
CofG_M.jann   ...........
CofH_M.jann   ...........
LipA_E.coli   ...........
MiaB_E.coli   ...........
NifB_A.vine   ...........
BssD_T.arom   ...........
NrdG_E.coli   ...........
NirJ_P.aeru   ...........
SplB_B.subt   ...........
```

**Figure IV.11: Overlay of the 4Fe-4S cluster and bound AdoMet from each of the AdoMet radical enzymes.** Colors are as follows: iron, ruby; sulfur, gold; oxygen, red; nitrogen, blue. In order to distinguish the 5 enzymes, the AdoMet carbons of each enzyme are colored as follows: teal, PflAE; magenta, HemN; orange, BioB; green, MoaA; purple, LAM.
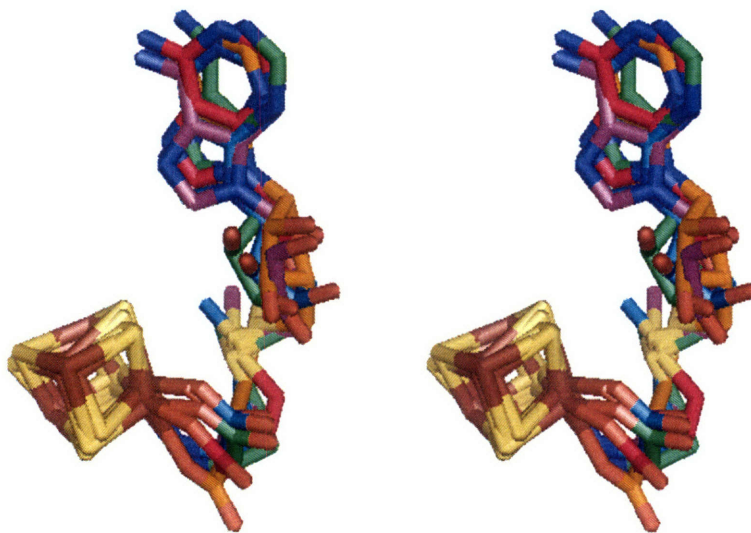
**Figure IV.12: Stereoviews of each AdoMet radical enzyme AdoMet binding site.**
The protein backbone is shown as grey cartoons, with AdoMet and the 4Fe-4S cluster
shown in sticks (AdoMet, green carbons; iron, ruby; sulfur, gold) and core β strands
labeled 1 – 6. Protein sidechains that interact with AdoMet are shown as lines with
carbons colored dark grey. Hydrogen bonding contacts are shown as red (within 3.2 Å
distance), green (3.2 – 3.7 Å) or yellow (more than 3.7 Å) dashed lines. Shown in this
figure are the AdoMet binding sites of (a) PflAE (b) HemN (c) BioB (d) MoaA (e) LAM.
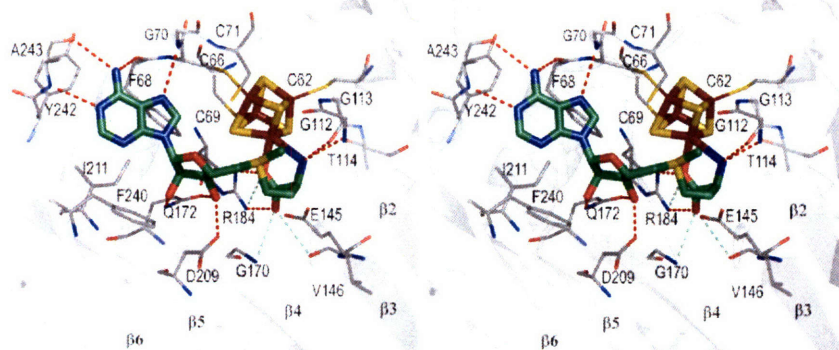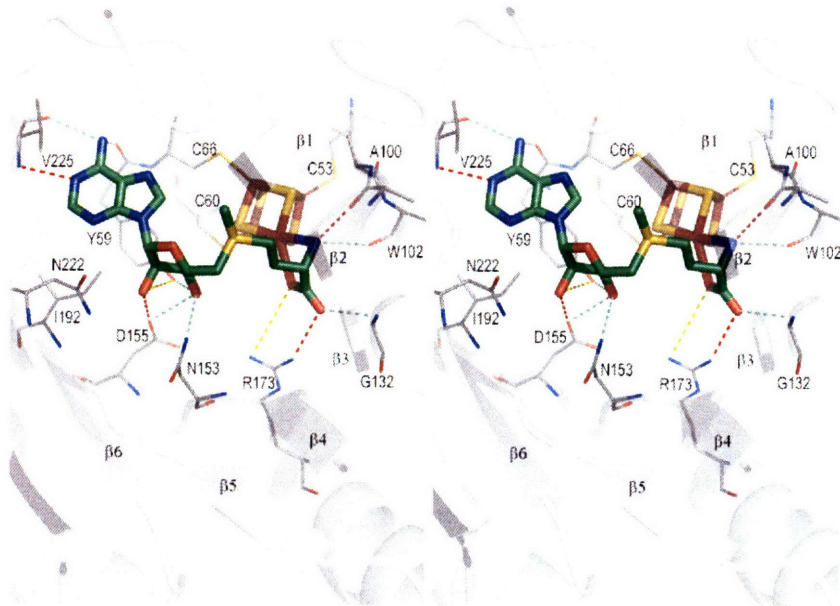


a  PflAE
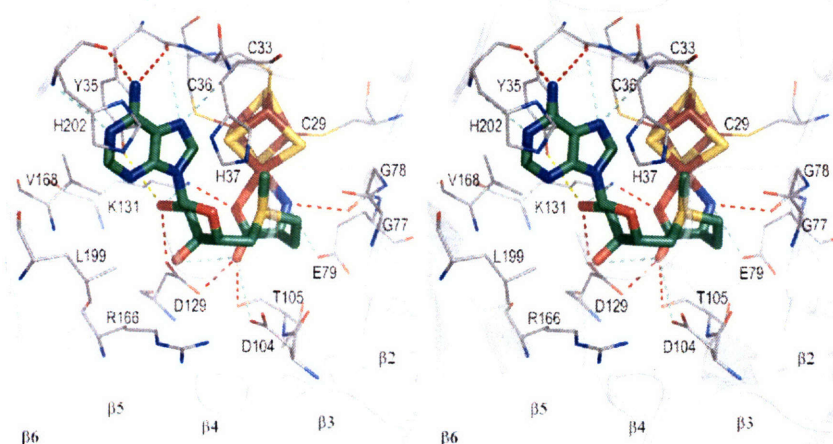


b  HemN

**Figure IV.12**, continued.

## c BioB



## d MoaA



184
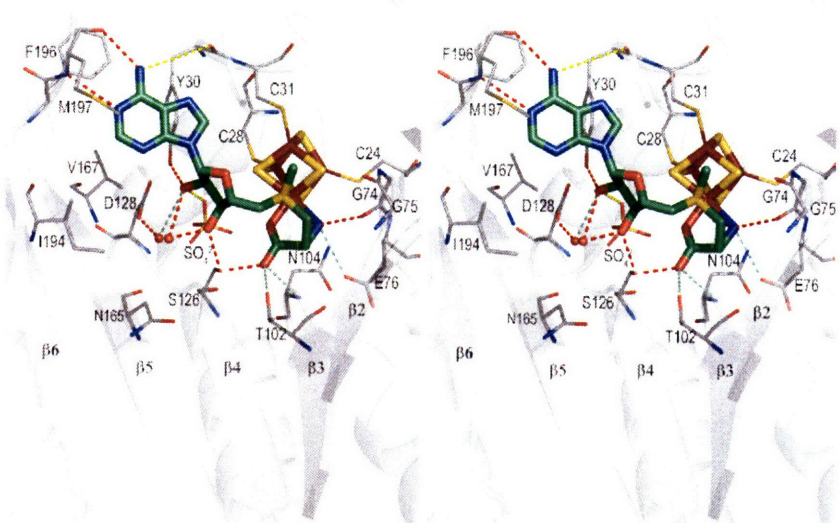
**Figure IV.12,** continued.

e LAM
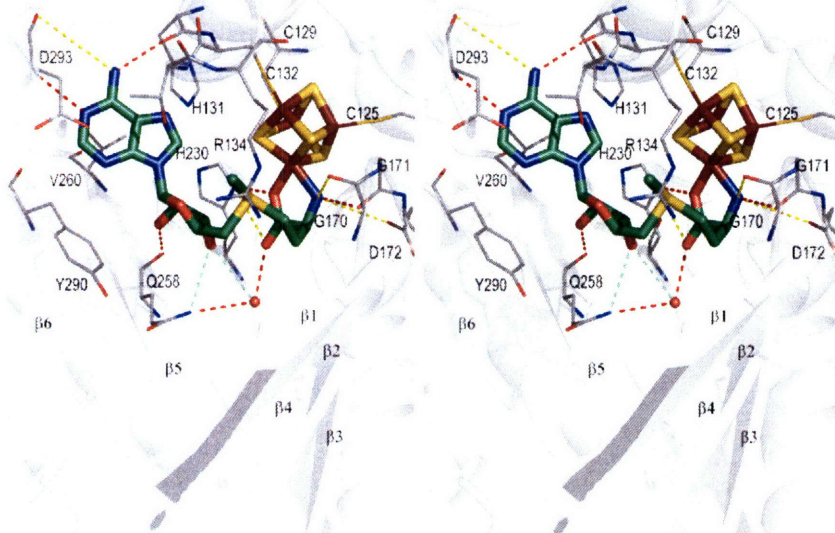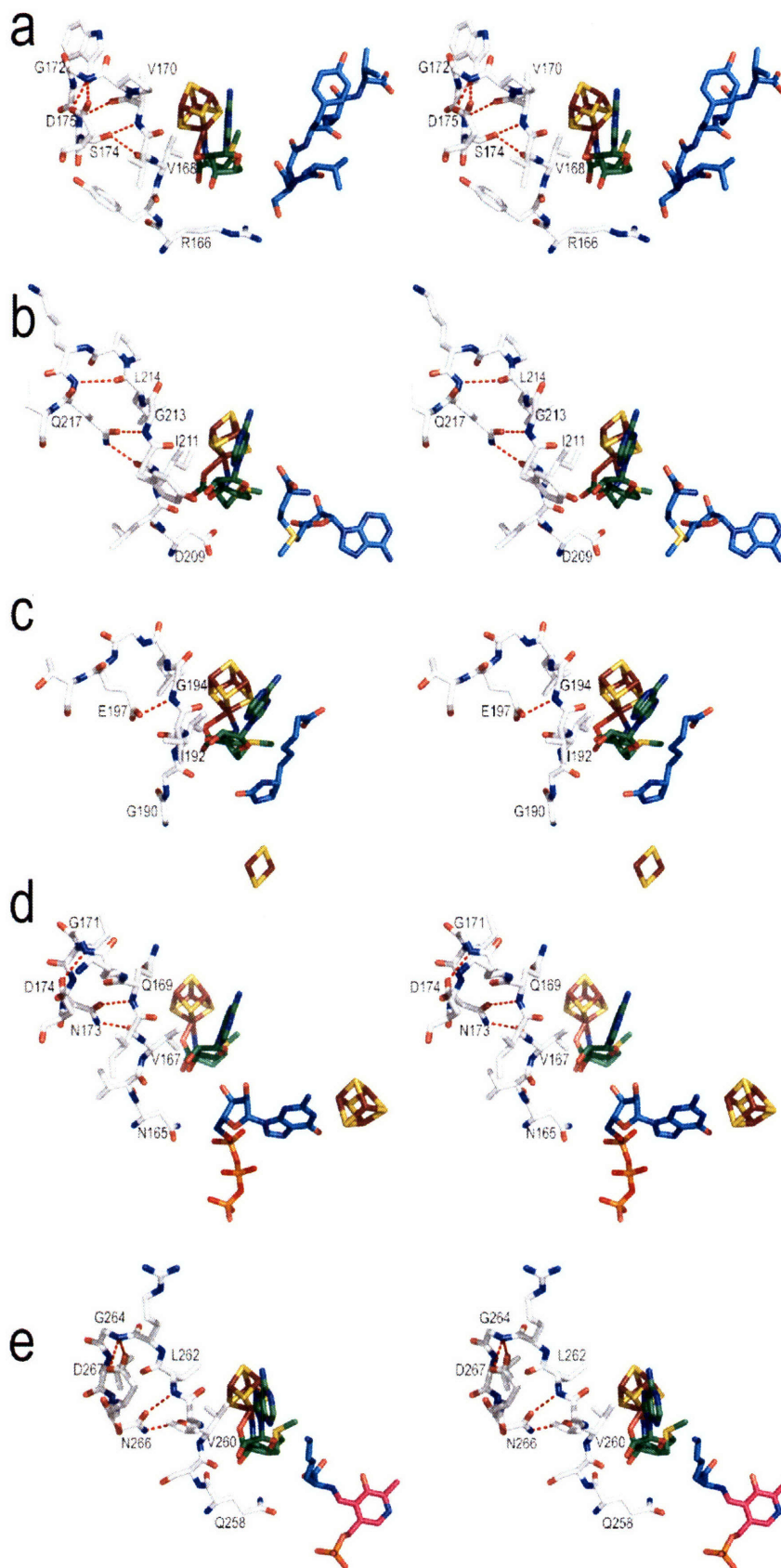
**Figure IV.13: The GxIxGxxE motif.** AdoMet and the 4Fe-4S cluster and the protein residues making up this motif are shown in sticks, colored as in Figure IV.4. Shown are (a) PflAE (b) HemN (c) BioB (d) MoaA (e) LAM. Note that, with the exception of BioB, the sidechain of the first residue of the motif is positioned similarly with respect to the AdoMet ribose and atom C5′. Each protein also has a similar sidechain-to-backbone hydrogen bond in the loop following β5 (shown as a red dashed lines). See text for more details.

**Figure IV.14: Stereoview of the typical MTase AdoMet-binding fold.** A representative MTase, catechol-O-methyltransferase (COMTase) )[66], is shown here in cartoons with strands colored yellow and helices colored teal. AdoMet, colored as in other figures, is observed bound at the C-terminal end of a mixed, mainly parallel β-sheet.
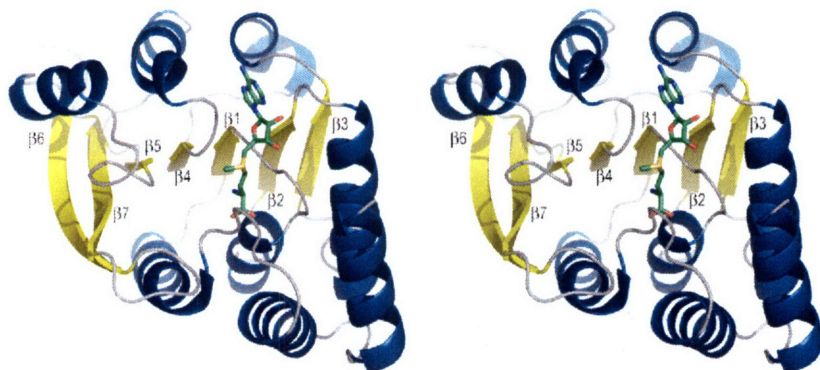
**Figure IV.15: Stereoview of other AdoMet binding folds**, all with AdoMet bound. (a) a typical SET domain (2NXE)[75], (b) AdoMet synthetase (2OBV)[76], (c) the MetJ repressor (1CMC)[78] and (d) the methionine synthase reactivation domain (1MSK)[79]. Each is shown in cartoon representation, colored as in Figure IV.14.
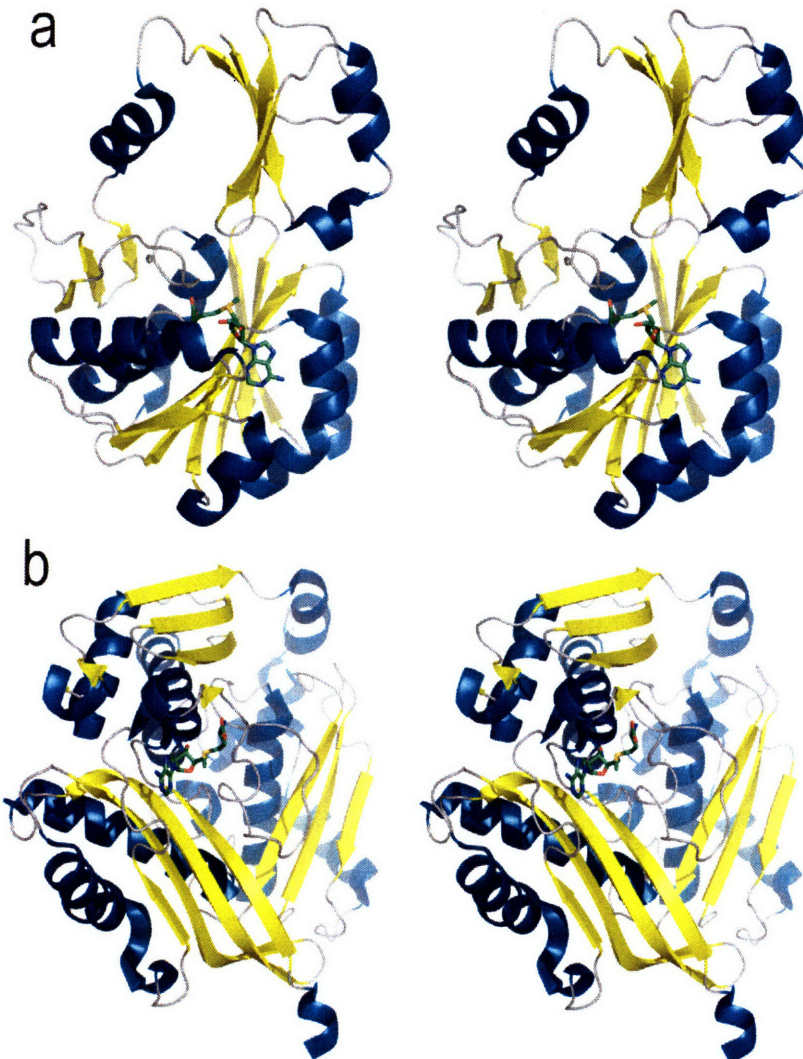
**Figure IV.15**, continued: (c) the MetJ repressor, (d) the methionine synthase reactivation domain.
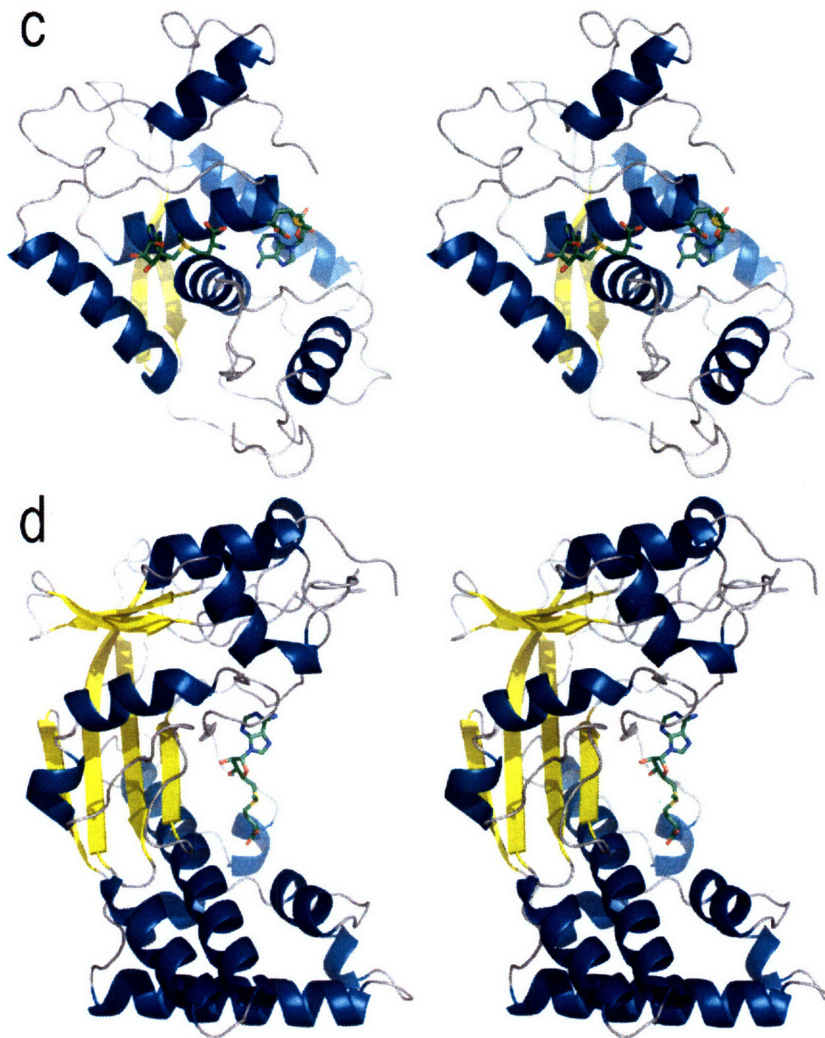
**Figure IV.16: The general substrate binding site in AdoMet radical enzymes**. The five structures were superimposed based on their AdoMet radical cores for this figure. Shown are the five "substrates" of the enzymes in sticks, colored as follows: teal, PflAE; magenta, HemN; orange, BioB; green, MoaA; purple, LAM. Also shown is the AdoMet radical core and 4Fe-4S cluster of PflAE only with β sheets labeled, as well as the AdoMet of PflAE in sticks, with carbons colored teal for clarity. The BioB 2Fe-2S and MoaA 4Fe-4S clusters are also shown, displayed as in other figures.
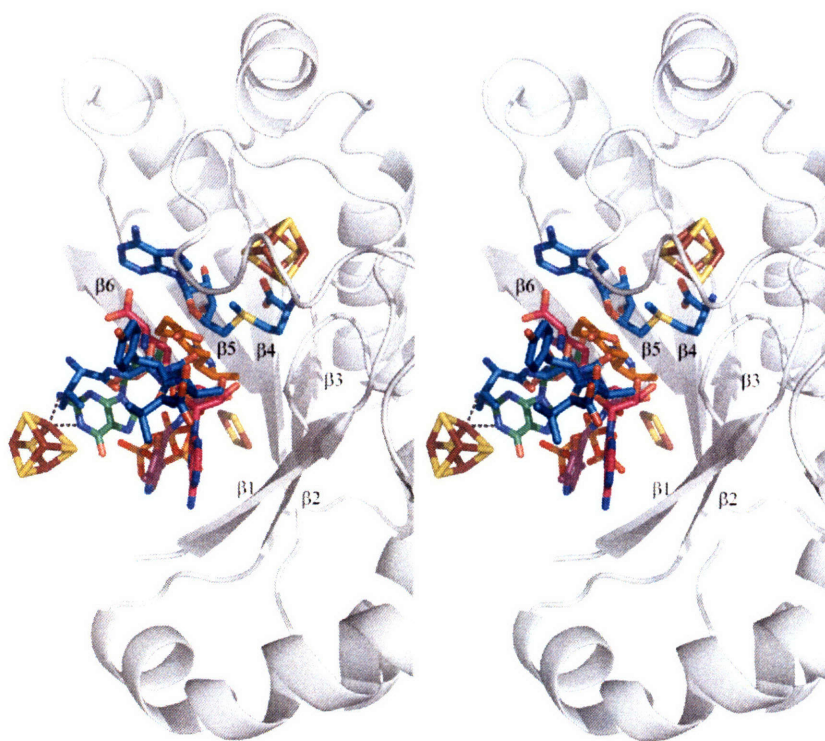
**Figure IV.17: The substrate binding site of each AdoMet radical enzyme**, highlighted to emphasize the elements of each protein involved in forming the binding site. Shown in this figure are the highlighted substrate binding sites of (a) PflAE with seven-mer peptide substrate (b) HemN with SAM2 (c) BioB with dethiobiotin and 2Fe-2S cluster(d) MoaA with putative substrate 5'-GTP and 4Fe-4S cluster (e) LAM with PLP and lysine. Each figure is shown in stereo and colored as in panels *a* and *b* of Figures IV.4 – IV.8 with the exception that the cluster binding loop is not colored magenta, and the cartoon is colored red at the backbone atoms of residues that contact the substrate.

## a PflAE
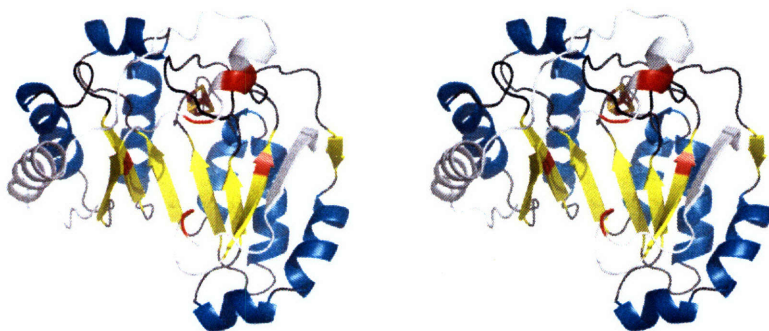


## b HemN
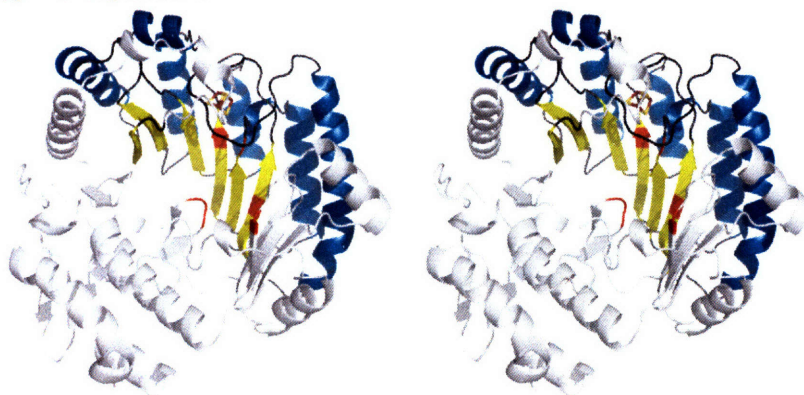


## c BioB
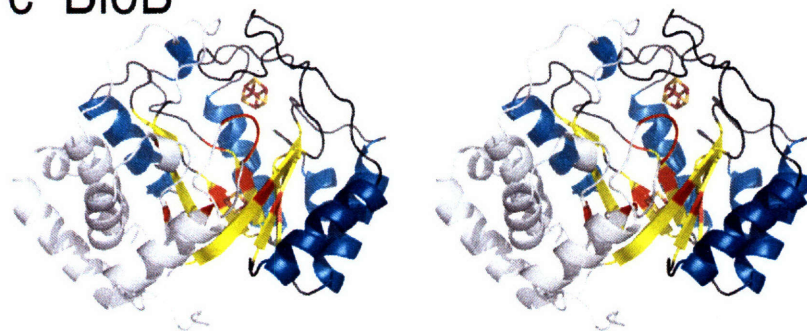
**Figure IV.17**, continued.

d MoaA
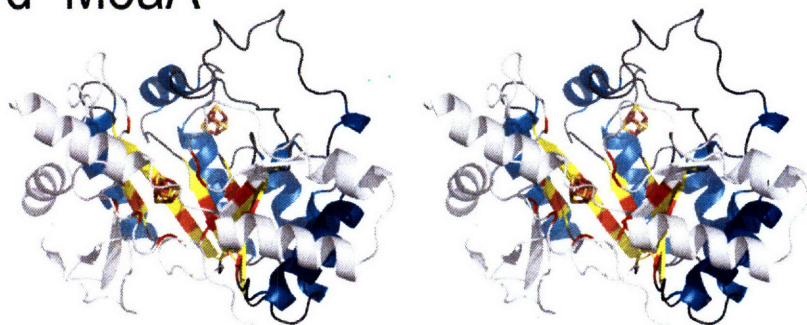


e LAM
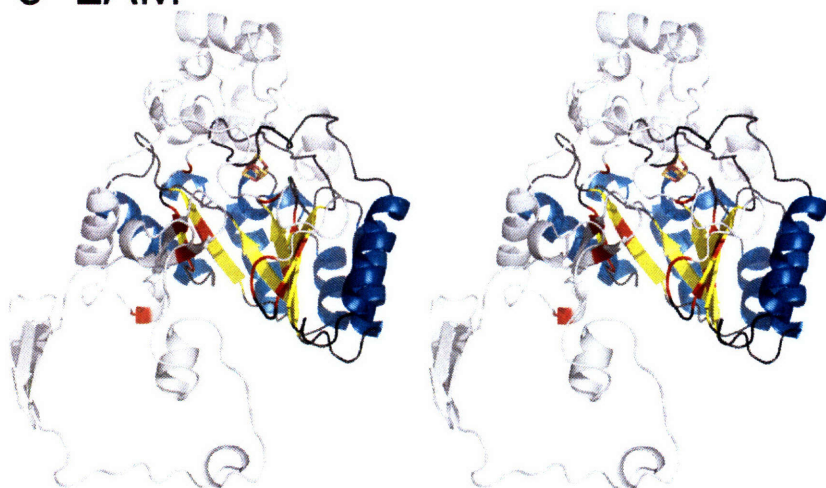
**Figure IV.18: The relative positioning of the substrate hydrogen atom abstraction point in the AdoMet radical enzymes.** This figure was generated from the five enzyme structures superimposed on their 4Fe-4S clusters and AdoMet only, to give a more accurate comparison of the relative positions of the substrates with respect to the AdoMet. The backbone, 4Fe-4S cluster and AdoMet are shown for the core of PflAE only, and colors are as described in Figure IV.16. C5′ of AdoMet is shown as a sphere. The atoms from which hydrogen abstraction is known to occur (i.e. Cα of $G_{734}$ of the peptide, C6 and C8 of dethiobiotin, and Cβ of lysine) are shown as spheres. In the case of MoaA, because AdoMet was not bound to the structure with 5′-GTP, the AdoMet-bound structure (1TV8) was superimposed as described, and the 5′-GTP-bound structure (2FB3) was then superimposed on that MoaA model.
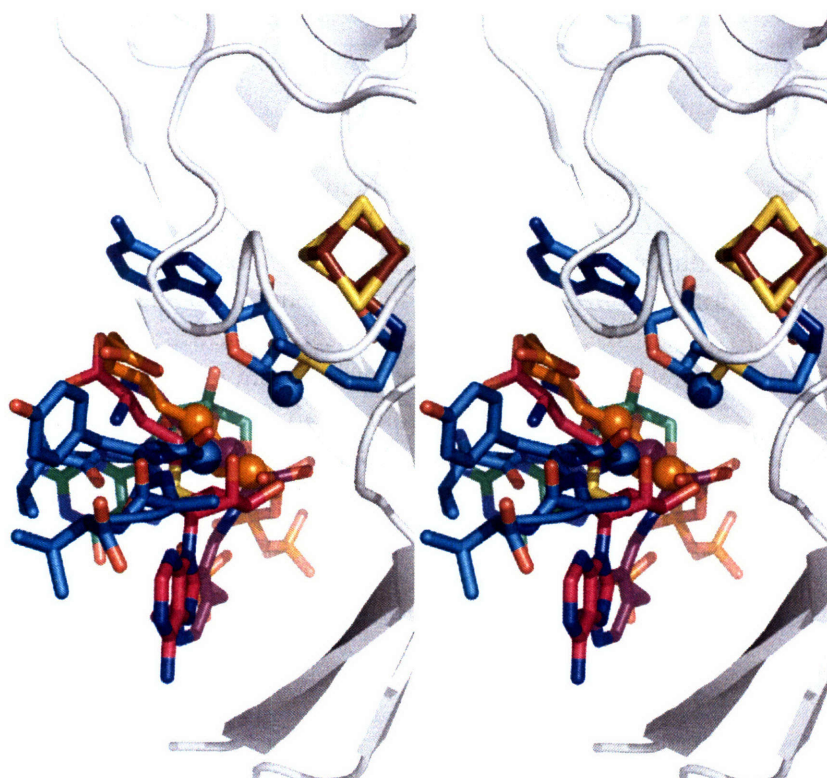
**Figure IV.19: Stereoview of substrate binding by PflAE**. The 4Fe-4S cluster, AdoMet and substrate are displayed as in Figure IV.4.a. PflAE sidechains that interact with the substrate are shown as lines, colored as in Figure IV.12.
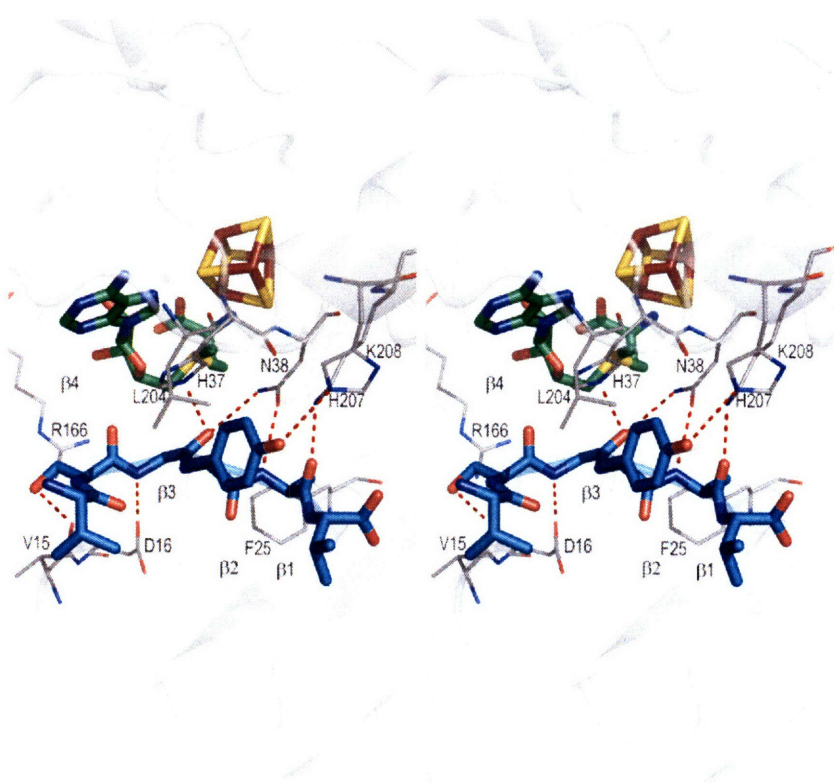
**Figure IV.20:  Stereoview of SAM2 binding site in HemN**.  The 4Fe-4S cluster, AdoMet and SAM2 are displayed as in Figure IV.5.a.  HemN sidechains that interact with SAM2 are shown as lines, colored as in Figure IV.12.  The β finger motif and β strand from the C-terminus that extends the AdoMet radical core region are both shown in light pink (see text).
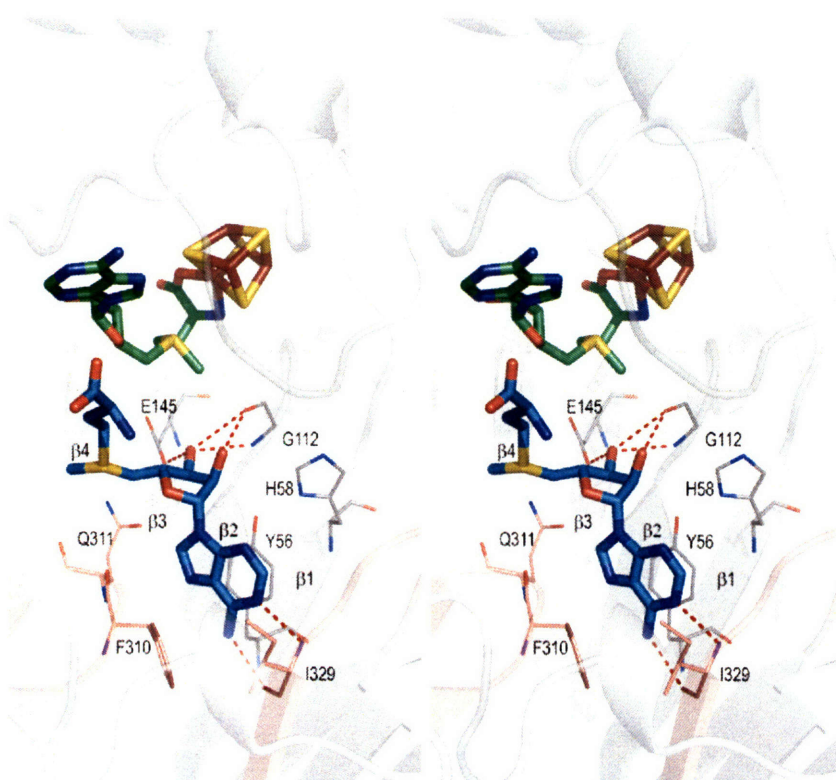
**Figure IV.21: Stereoview of (a) dethiobiotin binding and (b) the 2Fe-2S cluster of BioB.** The 4Fe-4S cluster, AdoMet and dethiobiotin are displayed as in Figure IV.6.a. BioB sidechains that interact with the substrate are shown as lines, colored as in Figure IV.12.
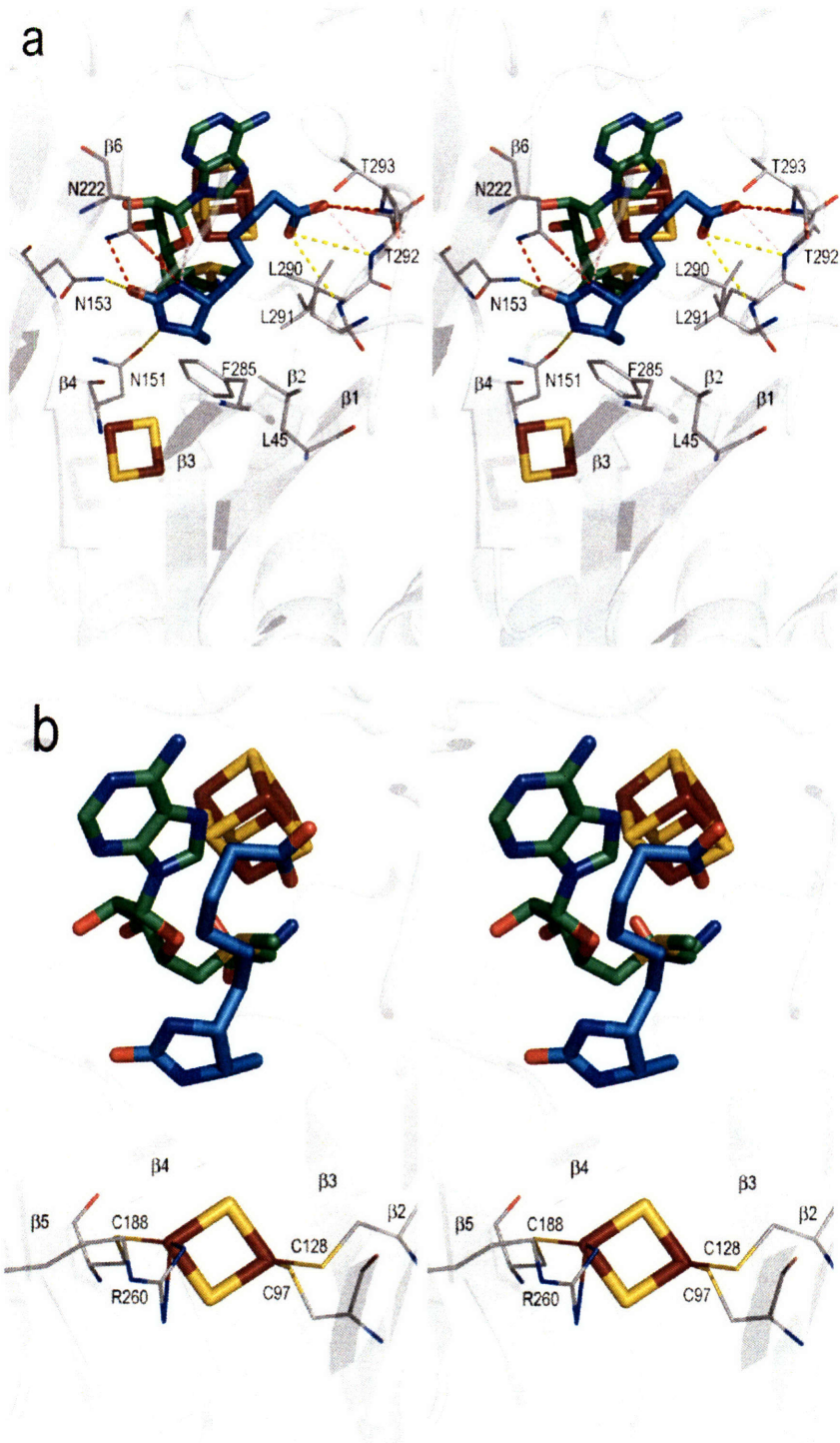
**Figure IV.22: Stereoviews of (a) 5′-GTP binding (b) the second 4Fe-4S cluster of MoaA.** The 4Fe-4S clusters, AdoMet cleavage products and GTP are displayed as in Figure IV.7.a. MoaA sidechains that interact with GTP are shown as lines, colored as in Figure IV.12.
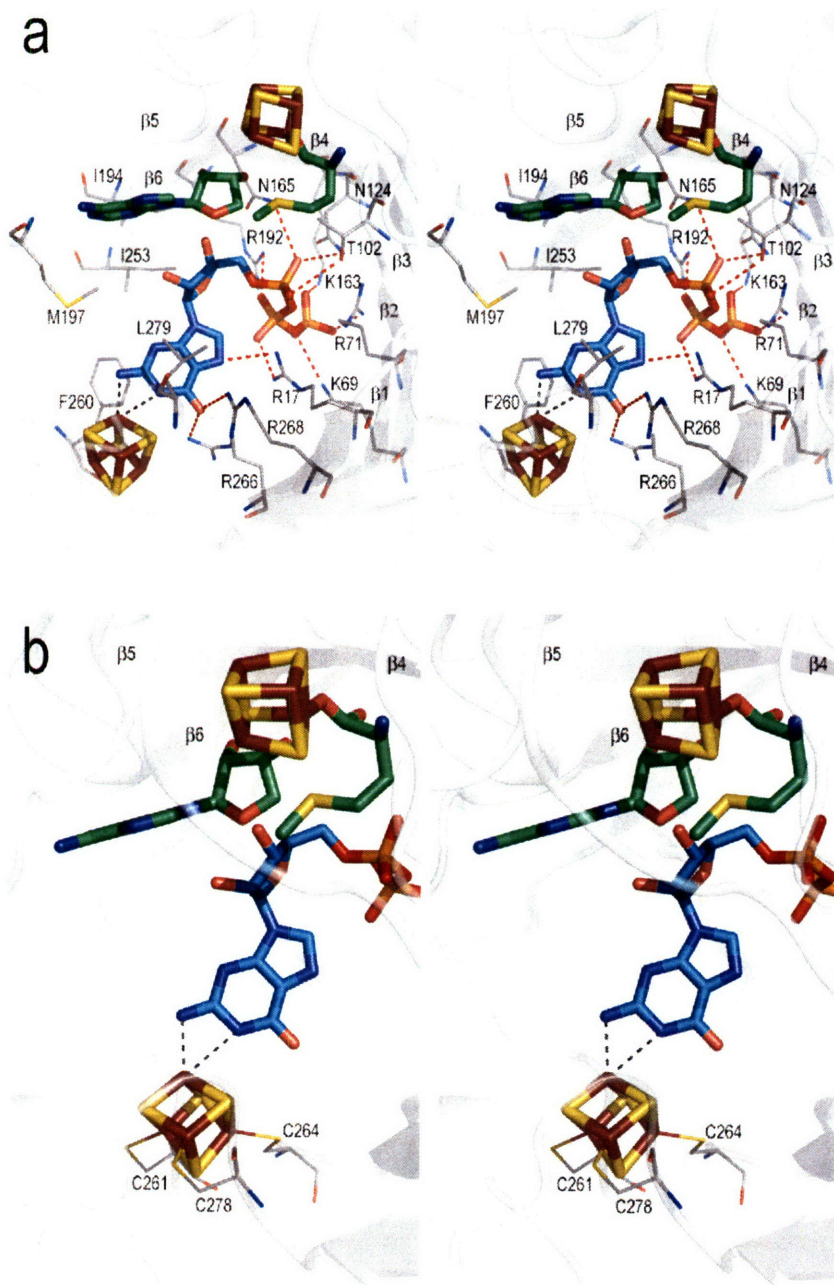
**Figure IV.23: Stereoview of (a) substrate and (b) cofactor binding by LAM.** The 4Fe-4S cluster, AdoMet, Lys substrate molecule and PLP are displayed as in Figure IV.8.a. LAM sidechains that interact with the substrate are shown as lines, colored as in Figure IV.12. In (b), part of a second molecule is shown in light blue.
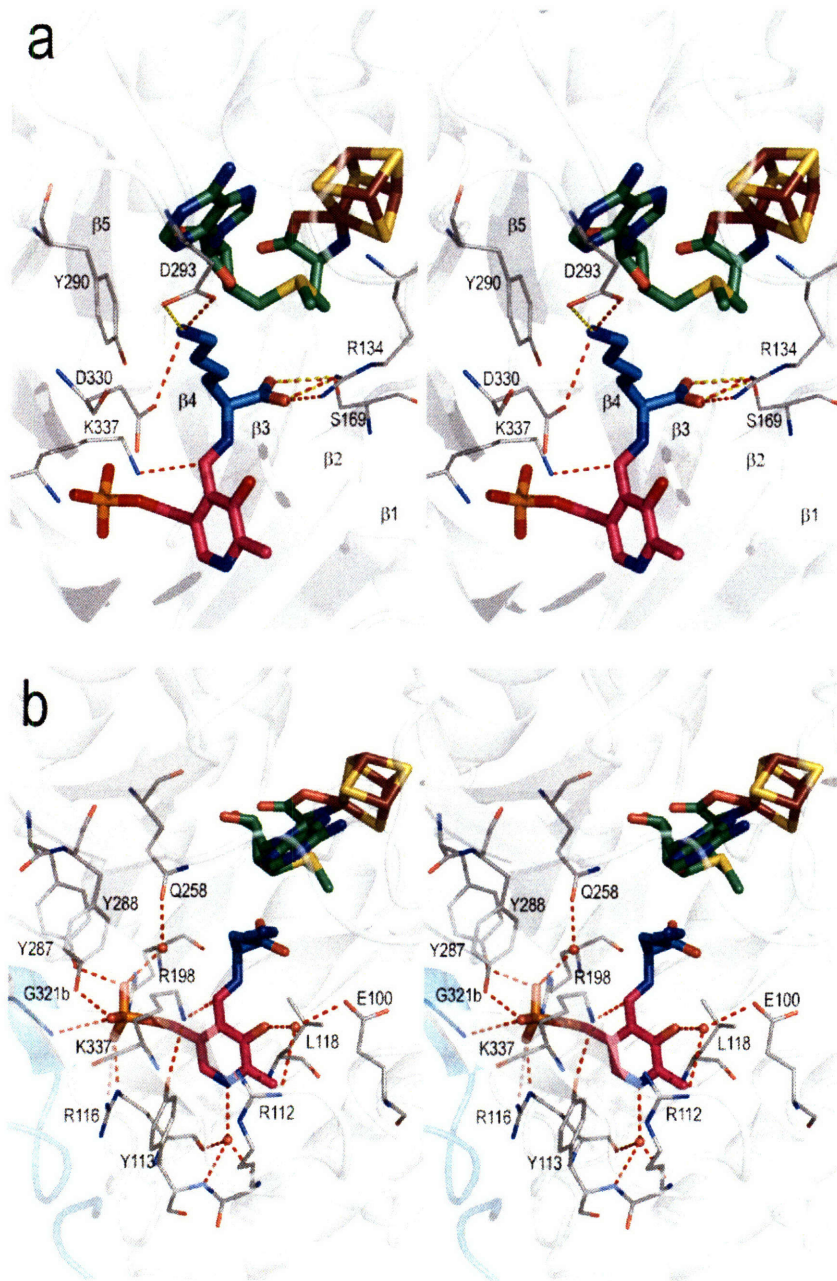
**Figure IV.24: Putative physiological reductant interaction surface for the AdoMet radical enzymes**. Shown in stereoview are the surfaces and cartoon representation of (a) PflAE, (b) HemN, (c) BioB, (d) MoaA and (e) LAM from the back of the enzyme, meaning the opposite side with respect to the lateral opening of the AdoMet radical core, roughly the same view as in Figure IV.9.b. The surface is shown at 50% transparency and is colored as a rainbow according to the extent of sequence conservation, with red being 100% conserved and blue as 0% conserved. The extent of conservation was calculated with ESPript, using a sequence alignment generated as follows. A BLAST search was conducted via the ExPASy proteomics server using the sequence of the enzyme that is structurally characterized as the query. The top 100 sequences were input for alignment by ClustalW. The alignment output by ClustalW was then used as input to ESPript.

a PflAE
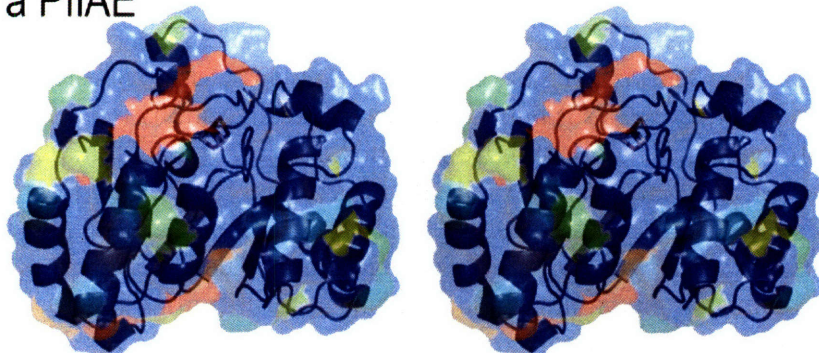


b HemN
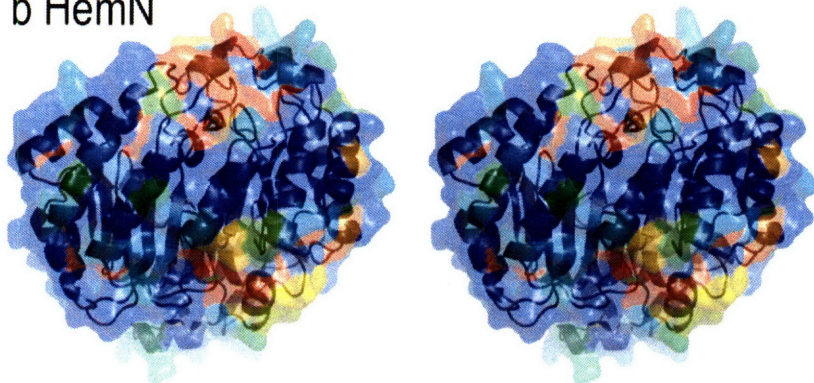
**Figure IV.24**, continued.

c  BioB
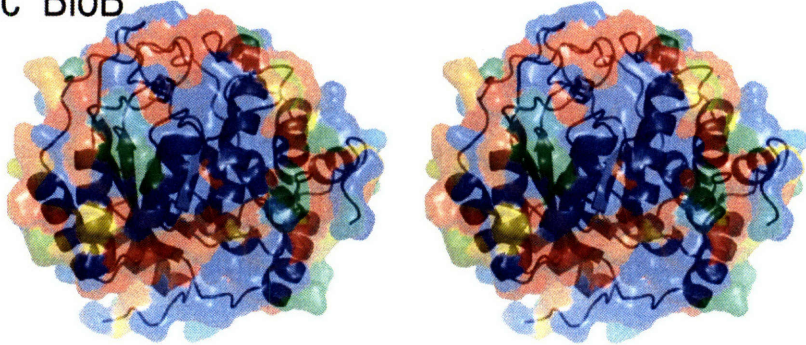


d  MoaA
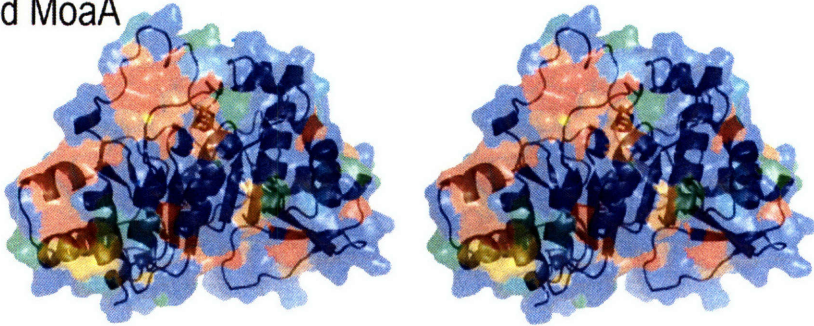


e  LAM
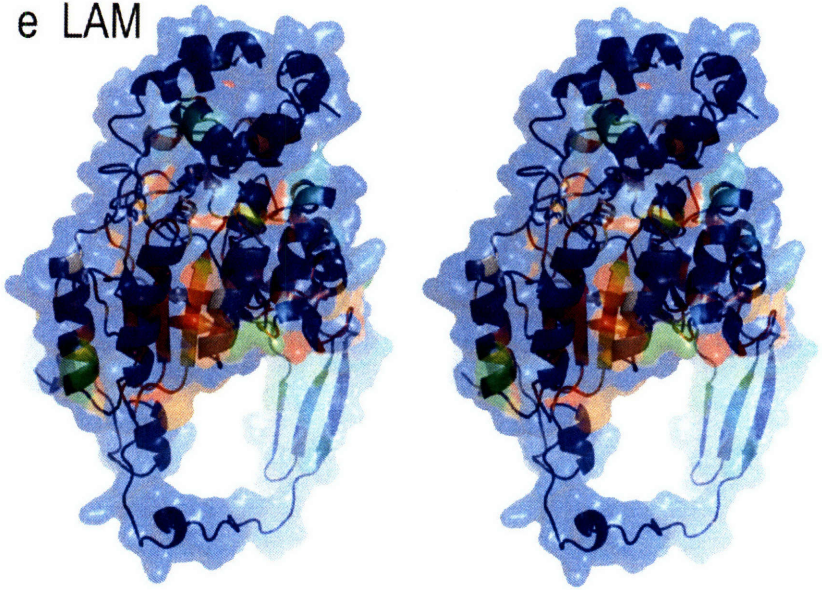
# Chapter V: Appendices

## Appendix 1: Further studies of iron-sulfur cluster biosynthesis

### A1.I Summary

In eukaryotes, the mitochondria are the site of the majority of iron sulfur cluster bioassembly. The primary machinery involved is the Isc gene cluster, which is highly conserved throughout the kingdoms of life. A second operon, the Suf gene cluster, has been implicated in FeS cluster synthesis under oxidative conditions. Researchers have postulated that the Suf operon acts as the primary biosynthetic machinery in certain types of cyanobacteria. The Suf operon may represent an evolutionary precursor to the Isc operon, and as it does not encode molecular chaperones (the Isc operon does), it may be a more viable crystallization target.

Our lab previously solved the structure of SufS from Synechocystis PCC sp 6803 by molecular replacement (see Chapter 2). This protein is a sulfur donor for iron-sulfur cluster biosynthesis. The biochemical characterization of that enzyme led to the hypothesis that another enzyme, identified by others as SufE, would be required for stimulating SufS cysteine desulfurase activity. SufE was shown to enhance the cysteine desulfurase activity of SufS by greater than 40-fold, perhaps by facilitating regeneration of the active site cysteine from the persulfide form through intermolecular sulfur transfer[1]. We became interested in the mechanism of this enhancement, and whether it could be investigated by X-ray crystallography. The next protein in the iron-sulfur cluster biosynthetic pathway is termed the "scaffold protein", denoted IscU, IscA, or SufA. Structural characterization by Cowan's group has led them to classify the *T. Maritima* IscU as a molten globule-like protein. Nuclear magnetic resonance, near-UV circular dichroism, 1-anilino-8-naphtalenesulfonic acid binding, free energy of unfolding, and hydrodynamic radius measurements collected on IscU are unclear, and consistent with either a molten globule or fully folded state[2,3]. They introduce the term MDC – multiple discrete conformers – to describe their results, and propose that the protein has a high degree of tertiary structure but experiences mobility on the μs to ms time scale.

The goals of the research described here was to crystallize several components of an iron-sulfur cluster biosynthetic pathway and to further structurally characterize those components, with the main objective of gaining insight into the Suf bioassembly system. Crystallization experiments were conducted focusing on the complex of SufS with the stimulatory protein SufE. Homologous proteins from three different organisms – *Synechocystis*, *Synechococcus*, and humans – were used for crystallization trials. It is also hoped that in the presence of an interacting partner – a cysteine desulfurase, for example – the scaffold protein will become more ordered, if there is indeed substantial mobility.

### A1.II. Crystallization experiments: SufS/SufE complex

The *E. coli* SufE and SufS proteins were received from S. Ollagnier-Choudens (Fontecave laboratory, CEA, Grenoble, France; buffer=50mM tris pH 7.5, 20mM NaCl), and the *Synechocystis* proteins were provided by B. Tirupati (Bollinger laboratory, Pennsylvania State University; buffer= 100mM Hepes pH 7.8, 10% glycerol). Samples

used for crystallization were the following: 1:1 ecSufES (two preps, mixed by Sandrine), 1:1 ecSufES+cys, 4:1 ecSufES, 1:1 sySufES.

Hampton Crystallization screens (primarily Crystal screens 1 and 2, Crystal screen lite, Index screen, Peg/Ion screen, Sodium malonate screen, and Peg/LiCl screen) were set up with the SufES complexes (for the best hit, see Figure A1.1).

Dynamic Light Scattering and Ni-NTA chromatography were used to determine whether the two proteins interact tightly. The proteins coeluted from a NiNTA column (200mM imidazole), confirming that their interaction is reasonably tight;
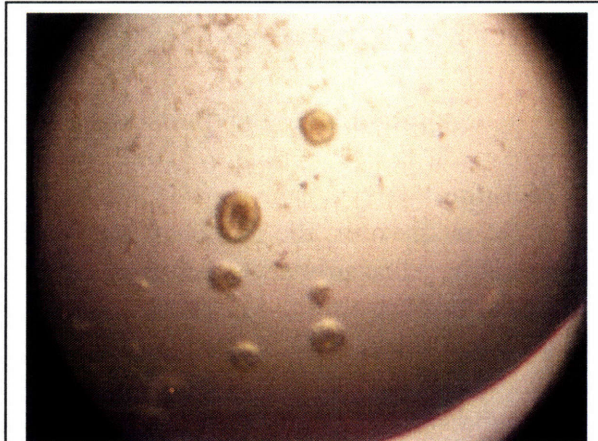


Figure A1.1: This phase separation was obtained in high salt conditions, and was the most promising hit obtained for the SufES complex. However, a gel showed that SufS alone comprised the round gel-like species depicted in this photograph, while both proteins are present in the drop.

however, the DLS results were more difficult to interpret. The % polydispersity of most samples was too high to allow us to draw conclusions confidently. The protein (SufE) is most likely not very stable under its storage conditions (50mM tris pH 7.5, 20mM NaCl), a conclusion supported by observations of precipitation after each freeze/thaw cycle. Storage conditions for the complex need to be optimized before crystallization trials can be resumed.


**A1.III. Crystallization experiments with bacterial scaffold proteins**

Two similar putative scaffold proteins, IscA from *Synechocystis* sp. PCC 6803 (syIscA) and IscA from *Synechococcus* PCC sp. 7002 (scIscA), were cloned into expression vectors and overproduced in *E. coli* and with the goal of crystallizing one of them, in complex with the cysteine desulfurase, if possible.


**A1.III.A. syIscA**

*A1.III.A.1 Construction of pSyIscAH$_6$*

A1.III.A.1a. Materials

*Synechocystis* sp. PCC 6803 genomic DNA was provided by B. Tirupati, Bollinger laboratory, Pennsylvania State University. All DNA oligos were obtained from IDTDNA technologies. Plasmid pET-15b was purchased from Novagen. DH5α and BL21 *E. coli* cells, as well as DNA gel standards, were purchased from Invitrogen. Mini Ready Gels were purchased from EmbiTec. The following kits from Qiagen were used: Miniprep kit, PCR Purification kit, and Gel Extraction kit. Restriction enzymes, shrimp alkaline phosphatase (SAP), T4 DNA ligase and T4 DNA kinase were obtained from New England Biolabs. Pfu Turbo was from Stratagene.

A1.III.A.1b. Methods and Results

The gene slr1417 was amplified by polymerase chain reaction (Figure A1.2) from *Synechocystis* genomic DNA using the following primers:
5' end, 5'-GGAATTC**CATATG**AGCCAAGCCACCGCTACCC-3'.
3' end, 5'-GC**GGATCC**TTAAACCCCAAAGGATTTACCAC-5'
Restriction endonuclease sites are marked bold (5' end, Nde1, 3' end, BamH1), and the gene is underlined.
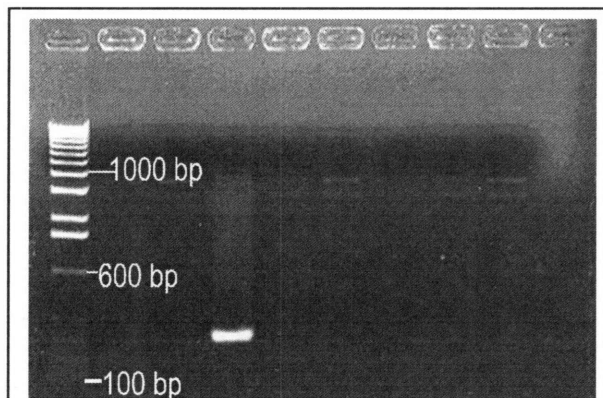


Figure A1.2: Test PCR results.
Lane 1, 100bp DNA ladder (Invitrogen).
Lanes 2-10, PCR reactions.

The target vector, pET-15b (Figure A1.3), was transformed into DH5α *E. Coli* cells for propagation and prepared from the organism using Qiagen's Miniprep Kit. The purified vector and amplified gene were digested with restriction enzymes BamH1 and Nde1. The vector was also treated with shrimp (SAP) to dephosphorylate the 5' ends of the digested vector. After digestion, the vector and insert were purified using Qiagen's PCR purification kit and ligated using T4 DNA ligase. After a heat denature step, the ligation reaction was transformed directly, without additional purification, into DH5α *E. coli* cells and plated out onto LB-agarose plates with ampicillin as the selection factor. The ligation yielded a very high background, evidenced by an equivalent number of colonies on vector-only control plates as on ligation reaction plates. Colonies were tested by PCR for presence of the insert (Figure A1.2). Plasmid testing positive for the insert was sequenced using the T7 promotor primer at the UPenn DNA Sequencing Facility. Alignment of the pSyIscAH$_6$ sequence with that of the slr1417 DNA sequence yielded 100% identity. pSyIscAH6 was then transformed into DH5α *E. coli* cells for propagation and BL21 *E. coli* cells for protein expression. The translated pSyIscAH$_6$ sequence is as follows, with the slr1417 sequence in bold:

MGSSHHHHHHSSGLVPRGSH**MSQATATQAKGIQLSDAALKHLLALKEQQGK
DLCLRVGVRQGGCSGMSYMMDFEEPNRATEHDEVFDYEGFQIICDRKSLLY
LYGLMLDYSNALIGGGFQFTNPNANQTCGCGKSFGV**

Figure A1.3: Vector map of pET-15b from Novagen.

*A1.III.A.2 Protein expression, purification and characterization*

A1.III.A.2a. Expression

Ampicillin at 100µg/mL was used at all times as the selection factor. BL21 *E. coli* cells expressing pSyIscAH₆ were used to inoculate 10mL LB cultures, which were then allowed to grow at 37°C overnight. The next morning, 1L cultures were inoculated with the overnight cultures and grown to an $OD_{600}$ of 0.8 – 1. The cultures were then induced by addition of IPTG to 0.5-0.8 mM. Cells were harvested by centrifugation after 4-6 hours growth at 37°C and stored at -80°C.

A1.III.A.2b. Purification

Cells were resuspended in Buffer A (50mM sodium phosphate, 5mM DTT, 150mM NaCl, pH 7.8) and sonicated to lyse (30 seconds of 2 sec on / 1 sec off, cycled 6-8 times). The cell break supernatant was then incubated on prerinsed NiNTA resin (rinse buffer was Buffer A + 10mM imidazole) at 4°C for 1-2 hours. The resin was rinsed and then washed with Buffer A + 50mM imidazole, and protein was eluted with Buffer A + 150mM imidazole. After dialysis to remove the imidazole, protein was run over a DEAE anion exchange column equilibrated in Buffer B (50 mM tris, 5mM DTT, pH 7.8) and eluted with a sodium chloride salt gradient. Protein fractions were pooled, concentrated, dialyzed into Buffer B + 150mM NaCl, 5% glycerol, aliquoted and stored at -80°C.

Figure A1.4: syIscAH6 purification. Lane 1: Benchmark protein MW standards (Invitrogen - bands correspond to the following molecular weights in kDa: 5.3, 13.5, 19.0, 24.5, 36.6, 48.0, 62.4, 79.0, 110.2, 172.0). Lane2: CBP. Lane3: CBS. Lane4: NiNTA column flow-through. Lane5: 10mM imidazole rinse. Lane6: start of 50mM imidazole wash. Lane7: end of 50mM imidazole wash. Lanes8-15: 150mM imidazole elution. Lane 16: Protein ladder. Lanes 17-19: DEAE fractions, peak 1. Lanes 20-30: DEAE fractions, peak 2.

Thus far, each separate purification has resulted in syIscAH$_6$ (expected molecular weight = 12.9 kDa) with a contaminant that runs between the 24.5kDa and 36.6kDa molecular weight standards. This contaminant has not been successfully separated from syIscAH$_6$, so N-terminal microsequencing and MALDI mass spectrometry were used to identify it. Interestingly, the sequencing results suggest that this band also contains syIscAH$_6$, perhaps in a disulfide-bridged dimeric form.

*A1.III.A.2 Preliminary biochemical characterization*

Reconstitution of syIscAH$_6$ was achieved anaerobically by direct addition of $Fe^{2+}$ (in the form of ferrous ammonium sulfate or ferrous sulfate) and $S^{2-}$ (sodium sulfide) to a ~4mg/mL protein solution in the presence of DTT. A dark brown color and substantial precipitation formed immediately upon addition of $Fe^{2+}$ and $S^{2-}$, but the protein was incubated anaerobically for approximately 2 hours before use of a G25 purification step to remove small molecules. UV-Vis spectra confirm the presence of a cluster, but occupancy has not yet been determined.

Reconstitution of the cluster was also attempted through addition of $Fe^{2+}$ and cysteine in the presence of catalytic amounts of *Synechocystis* SufS. Upon anaerobic incubation (without the addition of DTT), the protein solution took on a slight pink color. Unfortunately, attempts to observe a protein-bound FeS cluster were unsuccessful, perhaps due to precipitation of syIscAH$_6$ in the presence of a high concentration of $Fe^{2+}$.

In order to determine whether there is a strong interaction between syIscAH$_6$ and sySufS, protein preparations with 1:1 and 2:1 molar ratios (syIscAH$_6$:sySufS) were chromatographed on size exclusion (Biosec SEC-S, Phenomenex) and NiNTA affinity (Qiagen NiNTA His-bind Resin) columns. The size exclusion column did not have the resolution capability required to distinguish between the proteins alone and in complex at the flow rates used. The NiNTA affinity chromatography experiments did not show any

interaction between the two proteins, as sySufS eluted with 10mM imidazole, and syIscAH$_6$ stuck to the column (eluted with 150mM imidazole).

## A1.III.B. scIscA

### A1.III.B.1. Purification of the untagged scIscA

Cells expressing the pScIscA plasmid (obtained from Gaozhong Shen, Bryant and Golbeck laboratories, Pennsylvania State University) were sonicated to lyse, after which the cell break supernatant was fractionated by ammonium sulfate precipitation steps. The pellet obtained after the 30-60% fractionation was redissolved and dialyzed, followed by an anion exchange (DEAE) chromatography step. To further purify scIscA, HIC (phenyl-G-sepharose), hydroxyapatite, and size exclusion chromatography were used. Because little amounts of purified protein were obtained with the established purification protocol, the gene was cloned into a histidine tagged plasmid.

### A1.III.B.2. Subcloning scIscA into pScIscAH$_6$

The gene encoding scIscA was amplified from pScIscA (provided by Gaozhong Shen, Pennsylvania State University) by PCR using the following primers:
5' end: 5'-GGAATTG**CATATG**GCACAAACAACTACTG-3' (NdeI cleavage site)
3' end: 5'-CG**GGATCC**TTAAACCCCAAAAGATTTC-3' (BamH1 cleavage site)
Restriction endonuclease sites are marked bold (5' end, NdeI, 3' end, BamH1), and the gene is underlined.



Figure A1.5: Test PCR, scIscA cloning
  Lane 1: 1kbp DNA ladder (Invitrogen)
  Lanes 2-10: Test PCR reactions
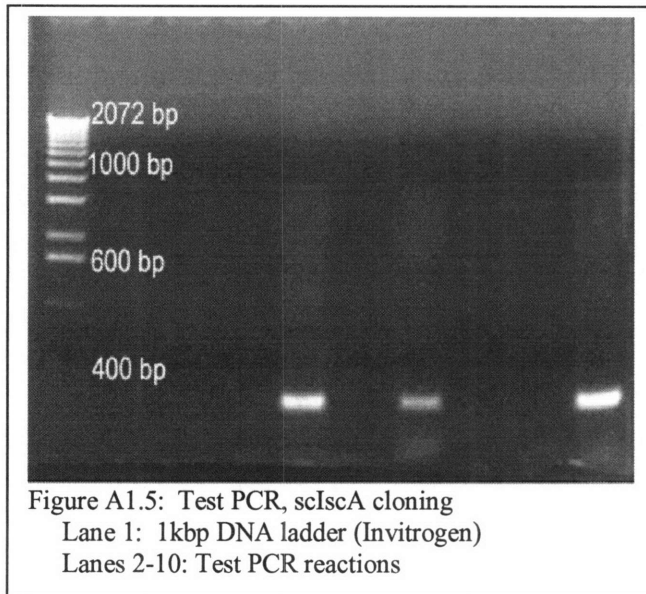
The target vector (pET-15b) and insert were, after purification, digested with NdeI and BamH1, and then ligated using T4 DNA ligase as described above. A high background was again observed on vector-only control plates (due to vector religation), and the PCR test was used again to identify successful ligation products. Plasmid encoding the scIscA gene was sent to the UPenn DNA Sequencing Facility to verify correct ligation, and then transformed into DH5α and BL21 *E. coli* cells for propagation and expression. scIscAH$_6$ was purified as described for syIscAH$_6$ in section A1.III.A.2b.

## A1.III.C. Tag removal from syIscAH$_6$ and scIscAH$_6$
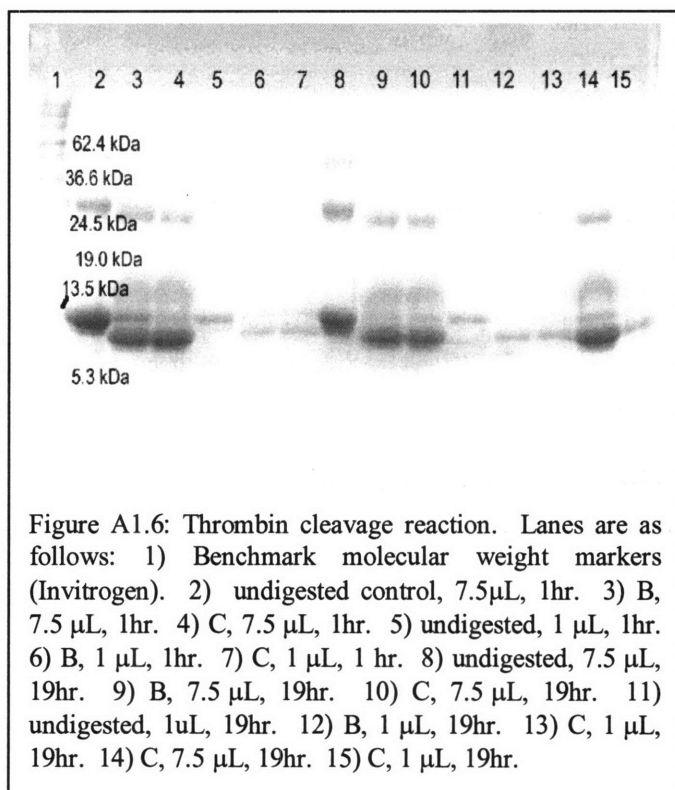


Figure A1.6: Thrombin cleavage reaction. Lanes are as follows: 1) Benchmark molecular weight markers (Invitrogen). 2) undigested control, 7.5µL, 1hr. 3) B, 7.5 µL, 1hr. 4) C, 7.5 µL, 1hr. 5) undigested, 1 µL, 1hr. 6) B, 1 µL, 1hr. 7) C, 1 µL, 1 hr. 8) undigested, 7.5 µL, 19hr. 9) B, 7.5 µL, 19hr. 10) C, 7.5 µL, 19hr. 11) undigested, 1uL, 19hr. 12) B, 1 µL, 19hr. 13) C, 1 µL, 19hr. 14) C, 7.5 µL, 19hr. 15) C, 1 µL, 19hr.

In order to help with crystallization and avoid adventitious iron binding, the N-terminal His tags of both syIscA and scIscA were removed. Thrombin cleavage of the His tag was carried out, confirmed by N-terminal sequencing, and optimized, and crystallization experiments were set up (with both apo and reconstituted syIscA) but did not yield any hits.

To determine the appropriate amount of incubation time and the amount of thrombin (Novagen) to add to the thrombin cleavage reaction, 300µg syIscA was mixed with (A) no thrombin, (B) 1U thrombin or (C) 5U thrombin dissolved in thrombin cleavage buffer (Figure A1.5). The reaction components were left mixing at 4°C for the times indicated. After further purification by size exclusion chromatography, samples were transferred to a nitrocellulose membrane and sequenced by the MIT Biopolymers Lab. The band expected to be product was confirmed as the thrombin-cleaved syIscA, with sequence GSHMSQATATQAKGIQ.

Typically, the cleavage reaction was left incubating overnight at 4°C. After completion of cleavage, the reaction was purified by size exclusion chromatography (Figure A1.6).

## A1.III.D. Crystallization of scaffold proteins

Crystallization experiments were set up using the Hampton Crystallization screens (primarily Crystal screens 1 and 2, Crystal Screen Lite, Index screen, Peg/Ion screen, Sodium malonate screen, and Peg/LiCl screen). The following protein samples were used in crystallization experiments: syIscAH$_6$, syIscA, reconstituted syIscAH$_6$ (anaerobic), reconstituted syIscA(anaerobic), scIscAH$_6$. No crystals or potential crystals were observed.

Crystallization experiments were also set up anaerobically using NFU (human IscU homolog) obtained from Wing Tong at NIH, but they also did not yield crystals.

Two structures of IscA were then published, both from *E. coli* (41% identity to slr1417)[4,5]. These papers outlined the tertiary fold of the *E. coli* scaffold protein and the quaternary structure as well. However, the active site cysteines thought to ligate the nascent cluster, which are located on the C-terminal tail, were disordered in both structures.



Figure A1.7: Size exclusion fractions following thrombin cleavage of syIscAH$_6$.
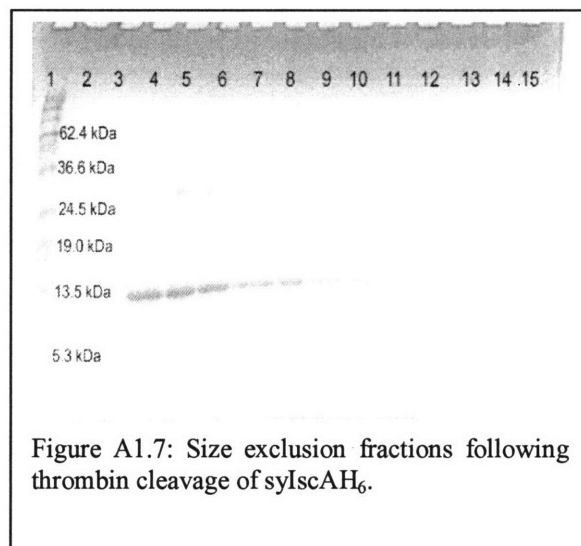
# References

1. Loiseau L, Ollagnier-de-Choudens S, Nachin L, Fontecave M, Barras F. *J. Biol. Chem.* (**2003**) 278:38352-9.

2. Mansy SS, Wu SP, Cowan JA. *J. Biol. Chem.* (**2004**) 279:10469-75.

3. Bertini I, Cowan JA, Del Bianco C, Luchinat C, Mansy SS. *J. Mol. Biol.* (**2003**) 331: 907-24.

4. Bilder PW, Ding H, Newcomer ME. *Biochemistry.* (**2004**) 43:133-9.

5. Cupp-Vickery JR, Silberg JJ, Ta DT, Vickery LE. *J. Mol. Biol.* (**2004**) 338:127-37.

# Appendix 2: Detailed Account of the Crystallographic Characterization of *E. coli* Pyruvate-formate lyase Activase

## A2.I. Summary

PflAE is an AdoMet radical enzyme in the activase subfamily, and catalyzes formation of a stable glycyl radical on its target enzyme, pyruvate-formate lyase.

Crystals of substrate-free PflAE in two macroscopic forms were obtained to eventually yield the problematic substrate-free structure: thin needles (10-20 µM X 200-1000 µM) and three dimensional crystals (50µM X 50µM).

The three-dimensional crystal form yielded data that were used to generate an initial substrate-free model. These crystals were reproduced through the use of high concentration salt in the wells of crystal trays, and the conditions were optimized to yield diffraction-quality crystals for data collection. Diffraction of these crystals was improved with detergents, and multiple datasets were collected at ALS and SSRL, as well as on our home X-ray generator. The structure was solved by anomalous dispersion techniques using the iron-sulfur cluster as the anomalous scatterer, and the current model comprises residues 5 to approximately 233, with two chain breaks. Refinement proved difficult in this case, with the Rfactors remaining high (specific values depend on the method of refinement and space group used).

A second protein sample, this one containing the seven-residue substrate (RVSGYAV) yielded a second crystal form that, upon increasing their size with detergents, was taken to SSRL for data collection. These data were of much higher quality than data from the substrate-free form, and yielded a model that refined well. This new model was then used to identify incorrect regions of the substrate-free form, eventually yielding another refined model.

This appendix contains a more detailed account of the crystallographic studies of the two forms of PflAE described in Chapter 3. In all of this work, solutions were degassed by bubbling with argon for 15 minutes per 2 mL of solution. Solutions with higher amounts of PEG were degassed slightly longer.

## A2.II. The substrate-free crystal form

### A2.II.A. Protein crystallization of substrate-free PflAE

*E. coli* PflAE (MW=28073 Da) was obtained from Meng Li and Jian Yang (in the laboratory of Prof. Joan Broderick, who is now at Montana State University though Meng and Jian stayed at Michigan State University). Initial crystals were obtained in 24-well trays from a Hampton Screen Kit after 2-3 days by mixing 1µL protein (20mg/mL in 50 mM hepes, 200mM NaCl, 1mM DTT, pH 7.5) with 1µL crystallization solution (0.1M tris pH 8.5, 25% PEG



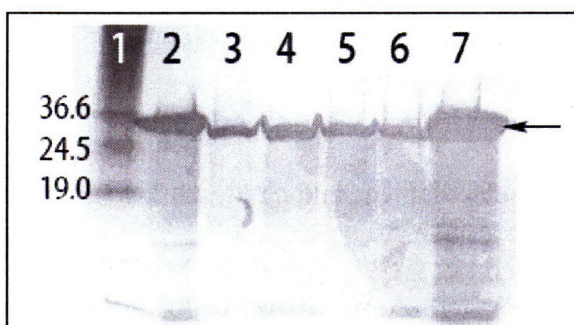Figure A2.1: The initial substrate-free PflAE crystal hit.

Figure A2.2: SDS-PAGE gel showing that the initial crystals contained PflAE. Lane 1, MW markers (molecular weights shown in kDa); 2, PflAE sample 1; 3, crystals A; 4, drop A; 5, crystals B; 6, drop B; 7, PflAE sample 2. An arrow denotes location of PflAE protein band.
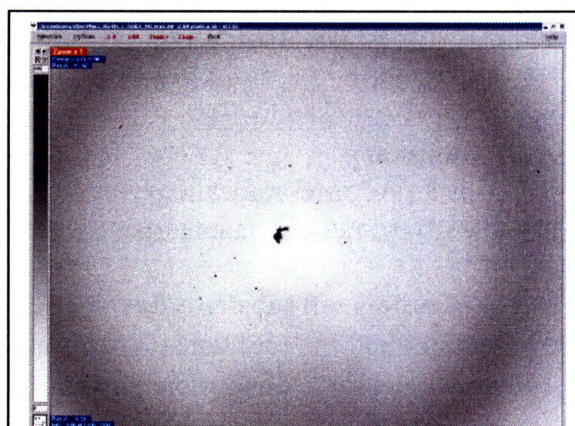


Figure A3: Diffraction from the initial substrate-free PflAE crystals.

3350), equilibrated against 1 mL 0.1M tris pH 8.5, 25% PEG 3350. These first crystals were very thin, brown needles with one very long dimension (~200-1000 microns – Figure A2.1). An SDS-PAGE gel confirmed that these crystals contained PflAE (Figure A2.2). The crystals diffracted to about 6 Å (Figure A2.3), and the cryoprotectants glycerol, MPD, PEG 400 and ethylene glycol were prepared to test at liquid nitrogen temperatures. Subsequent optimization of the two dimensional crystals did not yield better-diffracting crystals; however, some drops were later discovered to contain three-dimensional crystals. Efforts to reproduce and optimize these three-dimensional crystals continued.

Optimization of the crystallization conditions continued as follows. Additive and detergent screens (with conditions 0.1M tris pH 8.5, 25% PEG 3350) were set up in high throughput and Linbro formats. The rate of diffusion was increased by equilibrating the drops against 0.5 – 1.0 mL of a 2 – 2.5 M solution of ammonium sulfate, which eventually led to reproduction of the 3D crystals shown in Figure A2.4. It was later reasoned that faster dehydration of the drop somehow caused formation of the three dimensional crystals. More thorough screening of



Figure A2.4: The three-dimensional substrate-free crystal form.

conditions over wells containing ammonium sulfate followed, including microseeding, additive screening, and repeating the sparse matrix screening (Hampton Index Screen). The sparse matrix screen did yield similar 3D crystals in 5 crystallization conditions very close to the initial hit (see Table A2.1). Screens incorporating a wider pH range and 0.2M salt (NaCl, $MgCl_2$, $NH_4OAc$ and NaOAc) were also carried out, but yielded very few crystals (only 4 substantial crystals; other drops were empty or had very small, thin crystals).
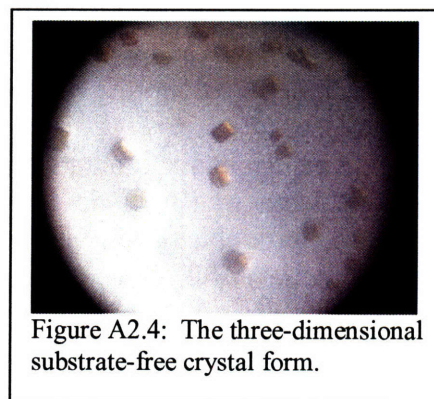
| |
|---|
| 0.1M tris pH 8.5, 25% PEG3350 |
| 0.1M bistris pH 6.5, 28% PEG MME 2000 |
| 0.2M NaCl, 0.1M bis-tris pH 6.5, 25% PEG 3350 |
| 0.2M NH4OAc, 0.1M bis-tris pH 6.5, 25% PEG 3350 |
| 0.2M NH4OAc, 0.1M hepes pH 7.5, 25% PEG 3350 |

Table A2.1: Substrate-free PflAE crystallization hits over high concentration ammonium sulfate wells.

The three dimensional crystals were used to screen cryoprotectant conditions. Five crystals were frozen and screened for each cryoprotectant, with PEG 400 and MPD yielding the best diffraction. Higher resolution reflections were streaky and diffuse in both cryoprotectants, and addition of Detergent #2 from Hampton's Detergent Screen 1 ($C_{12}E_8$) to the holding and cryo solutions improved this streakiness. Addition of the cryoprotectants (not carried out with PEG 400) to the crystallization conditions themselves did not yield quality crystals. Some experimentation was carried out to optimize the conditions including the detergent, including varying the drop volume and ratio, the well volume, and preincubation of the protein with the detergent. In all cases, the diffraction quality varies widely crystal-to-crystal, so a large number of crystals were screened to identify crystals good enough for data collection. The final cryoprotection solution used was 0.1 M Tris pH 8.5, 25% PEG 3350, 0.11 mM octaethylene glycol monododecyl ether and either 20% PEG 400 or 20% 2-methylpentane-2,4-diol (MPD).

### A2.II.B. Data collection and phasing of the substrate-free PflAE crystal forms

*A2.II.B.1. First home dataset – primitive tetragonal lattice*
The first PflAE dataset was collected at home on a crystal grown in a high throughput format tray with the following conditions: 0.1M tris pH 8.5, 26% PEG 3350. This crystal did not grow over ammonium sulfate – it formed over several months of evaporation, and grew off of the bottom of the HT tray. It was cryoprotected with 30% PEG 400, frozen, and later annealed into cryo solution for 30 seconds before data collection to improve the diffraction. A sample diffraction image is shown below. Data were collected to 3.5Å resolution, the crystal indexed with a primitive tetragonal lattice and scaled well as any of the P422 space groups. The cell dimensions were 58.9Å x 58.9Å x 134.0Å with angles 90° x 90° x 90°, and a reasonable value for solvent content (39.9%) was calculated with one molecule per asymmetric unit.

| | |
|---|---|
| Space Group | $P4_32_12$ |
| Resolution | 3.5 Å |
| Rsym | 10.7(20.9) % |
| % refl rejected | 0.20 % |
| Redundancy | 8.9 |
| I/sigma | 18.2(8.36) |
| % Complete | 99.6(98.4) |

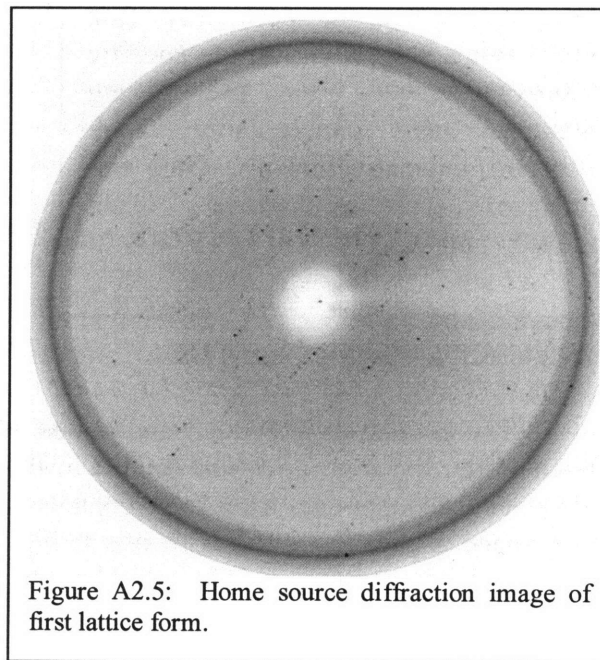Table A2.2: Data statistics from first home dataset of substrate-free PflAE.



Figure A2.5: Home source diffraction image of first lattice form.

211

Anomalous scattering from the 4Fe-4S cluster was then used to attempt to obtain phases from the home data (peak anomalous absorption for iron $\approx 1.7$ Å, copper $K\alpha$ radiation = 1.5418 Å. Output from Scalepack indicated a small amount of anomalous signal, and anomalous pattersons were calculated using XtalView. The Harker sections of those Patterson maps were examined to find one site corresponding to the iron sulfur cluster. A site (0.1957, 0.1581, 0.1137) was identified by hand in space group $P4_32_12$, but refinement and calculation of electron density maps in SHARP did not produce traceable maps (figure of merit (FOM) = 0.35, phasing power (PP) = 1.03 to 5.0Å; FOM = 0.18, PP = 0.51 to 3.5Å). Sites identified with SOLVE and CNS did not match Pattersons (all carried out in space group $P4_22_12$). Other attempts to use SOLVE or CNS, or direct observation of Pattersons, to find the site failed in the three space groups. Molecular replacement with the biotin synthase structure was also unsuccessful. In order to obtain a stronger anomalous signal, the crystal was shipped to SSRL for data collection, but died upon further annealing by the staff there.

The primitive tetragonal cell observed in the first dataset was never reproduced. All subsequent crystals were indexed as a primitive hexagonal space group.

*A2.II.B.2. Second home dataset – primitive hexagonal lattice*

The crystal used for this dataset came from the condition 0.1M tris pH 8.5, 26% PEG 3350, 0.1mM $C_{12}E_8$ (Hampton Detergent Screen 1 #2), crystallized over wells containing 2.5 M ammonium sulfate. The crystal was cryoprotected with the crystallization condition and 20% MPD, and data were collected to 3.5Å. The space group proved more elusive, with both centered orthorhombic (C2) and primitive hexagonal lattices a possibility (the latter having, depending on the specific reflections chosen for use during indexing, an unacceptably high distortion index at times). Unit cell dimensions were as follows: when indexed as C2, 100.3Å x 57.8Å x 116.2Å with angles 90°, 89.9°, 90°; when indexed as P3, 57.9Å x 57.9Å x 116.2Å with angles 90°, 90°, 120°.
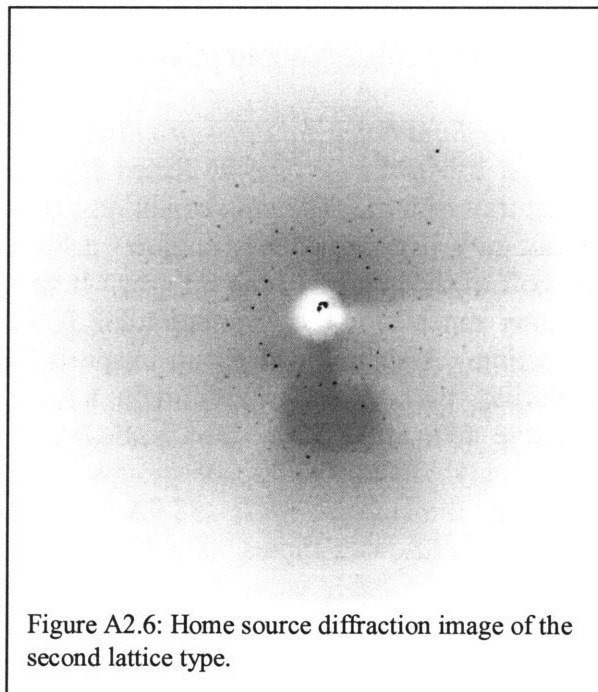


Figure A2.6: Home source diffraction image of the second lattice type.

The data were integrated and scaled with Denzo and Scalepack, and the results indicated the space group would be one of those based on P321. The mosaicity for this data and subsequent data from these crystals was extremely high (greater than 1.5°). For these space groups Z=6, and with one molecule per asymmetric unit the crystal would have a solvent content of 38%. Scalepack output files indicated a small anomalous signal, which was used to calculate anomalous Pattersons. The Pattersons were analyzed

by hand and a site was identified, but no useable electron density maps were obtained from this data. Crystals were taken to the synchrotron to collect higher resolution data.

## A2.II.B.3. Data from SSRL

A MAD experiment and two native datasets were collected at SSRL in March 2005. The crystals collected on were formed from 0.1M tris pH 8.6, 30% PEG 3350 in hanging drops over 2.5 M ammonium sulfate, and were cryoprotected with the same conditions containing 20% MPD. The unit cell and space group were similar to those of the second home dataset (primitive hexagonal, probably P321 with unit cell ~ 57 Å x 57 Å x 117 Å, 90° 90° 120°).

| Dataset | λ (A) | E(eV) | f' | f" | dist (mm) | exp time |
|---|---|---|---|---|---|---|
| peak | 1.73958 | 7127.9 | -6.9 | 3.9909 | 190 | 30 sec |
| inflection | 1.74166 | 7118.5 | -8.2593 | 2.3 | 190 | 20 sec |
| remote | 1.34761 | 9200 | | | 350 | 20 sec |
| native | 1.00089 | 12387 | --- | --- | 350 | 20 sec |
| native2 | 0.9797 | 12387 | --- | --- | 400 | 60 sec |

Table A2.3: Data collection information, SSRL.

Processing this data was tricky, because the $\chi^2$ values and mosaicity were very high. To help lower the $\chi^2$ values, the data was integrated with different mosaicity values and spot sizes, partials were added over different blocks of data to increase the number of legitimate observations, and the scale and B factors were fit during scaling. $P3_121$ and $P3_221$ were the space groups used at first for these datasets. Because of significant ice rings, the native2 dataset was not fully processed (Table A2.3, A2.4).

| SSRL | peak | inflection | remote | native1 | native2 |
|---|---|---|---|---|---|
| Beamline | 9.2 | 9.2 | 9.2 | 9.2 | 9.1 |
| Space Group | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ |
| Resolution (Å) | 2.8 | 3.9 | 3.2 | 2.65 | |
| Rsym (%) | 9.5(34.0) | 8.7(39.9) | 7.2(39.6) | 7.3(35.2) | |
| %refl rejected | 0.64 | 0.13 | 0.14 | 0.11 | |
| Redundancy | 5.4 | 5.6 | 5.8 | 5 | |
| I/sigma | 13.6(3.6) | 17.4(3.45) | 19.7(3.4) | 16.2(4.1) | |
| % Complete | 99.3(99.0) | 99.5(100.0) | 99.4(99.9) | 96.6(97.1) | |
| a (Å) | 56.327 | 56.317 | 56.39 | 58.441 | |
| c (Å) | 117.356 | 118.166 | 117.925 | 117.367 | |

Table A2.4: Data collection statistics, SSRL.

The Scalepack $\chi^2$ test indicated presence of a strong anomalous signal in the peak dataset, and anomalous Pattersons were calculated. Dispersive difference Pattersons were also calculated using the remote and inflection datasets. Pattersons were inspected by hand to identify the site, and SOLVE and CNS were both used to search for the site as well (see Table A2.5). Sites identified in any of these three ways were checked against the Pattersons and, if valid, refined in SHARP using either the peak (SAD) or all three wavelength datasets (MAD). No traceable maps were produced from any of these sites,

213

and problems were encountered with SHARP – the program refined the occupancies of the site too high. See Table A2.5 for site coordinates and sample figure of merit and phasing power statistics.

| Sites from dataset | SG | x | y | z | FOM | PP | traceable map | occ/b problems ? |
|---|---|---|---|---|---|---|---|---|
| home dataset2 | P3₁21 | 0.276 | 0.122 | 0.499 | 0.42 | 2.4 | no | |
| home dataset2 | P3₁21 | 0.1638 | 0.768 | 0.008 | | | | |
| SSRL peak | P3₁21 | 0.594 | 0.780 | 0.160 | 0.38 | 2.37 | no | yes |
| ALS peak | P3₁21 | 0.834 | 0.232 | 0.008 | 0.37 | 2.354 | no | yes |
| ALS peak - MAD | P3₁21 | 0.166 | 0.767 | 0.008 | 0.18 | 0.9-1.6 | no | ? |
| ALS peak - peak | P3₁21 | 0.166 | 0.767 | 0.008 | 0.37 | 2.343 | no | yes |
| ALS peak | P3₂21 | 0.233 | 0.834 | 0.008 | 0.37 | 2.31 | no | yes |
| ALS mad | P3₂21 | 0.769 | 0.597 | 0.159 | 0.82 | 0.94-4.4 | no | yes? |
| ALS peak | P3₁ | 0.600 | 0.830 | 0.991 | 0.66 | 0.9-1.6 | yes | no |
| | | 0.767 | 0.172 | 0.308 | | | | no |
| SSRL2 peak | P3₁21 | 0.402 | 0.235 | 0.158 | 0.73 | 0.989-1.701 | yes | no |

Table A2.5: Summary of the different heavy atom sites obtained from the substrate-free crystal form data with phasing statistics.

### A2.II.B.4. Data from ALS

Because no traceable electron density maps could be produced from the home or SSRL data using any of the sites identified by hand or by SOLVE/CNS, more data were collected at ALS. It was hoped that the new data would be of better quality (and lower mosaicity), allowing identification of the correct site.

| Dataset | $\lambda$ (A) | E(eV) | f' | f" | dist (mm) | exp time |
|---|---|---|---|---|---|---|
| peak | 1.74037 | 7124 | -6.9 | 3.9 | 175 | 30 sec |
| inflection | 1.74231 | 7116 | -8 | 2.3 | 175 | 30 sec |
| remote | 1.4 | 8856 | -0.505 | 2.7 | 250 | 20 sec |
| native | 1 | 12387 | | | 350 | 30 sec |

Table A2.6: Data collection information, ALS.

The crystals collected on at ALS were grown from 0.1M tris pH 8.0, 26% PEG 3350 (hanging drop over 2.5 M ammonium sulfate) and cryoprotected with 20% MPD in the crystallization condition. The diffraction patterns could be processed with a primitive hexagonal lattice or a centered orthorhombic lattice, as other data from these crystals, and had a similar unit cell. All the data were processed in P3₁21 first, because the intensity pattern of the reflections seems to match that space group well.

| ALS | peak | inflection | remote | native |
|---|---|---|---|---|
| Beamline | 5.0.2 | 5.0.2 | 5.0.2 | 5.0.2 |
| Space Group | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ |
| Resolution (Å) | 2.6 | 2.75 | 2.75 | 2.7 |
| Rsym (%) | 9.1(34.6) | 8.5(32.8) | 8.8(30.2) | 7.9(31.2) |
| % refl rejected | 0.42 | 0.27 | 0.16 | 0.39 |
| Redundancy | 5 | 5.1 | 4.3 | 5.9 |
| I/sigma | 19.5 | 20.2 | 21.6 | 20.9 |
| % Complete | 99.5(97.3) | 99.8(98.8) | 98.9(95.0) | 95.3(95.3) |
| a (Å) | 57.817 | 57.928 | 57.994 | 57.547 |
| c (Å) | 116.525 | 116.649 | 116.653 | 115.725 |

Table A2.7: Data collection statistics, ALS.

The new data processed in $P3_121$ were used to refine against sites obtained previously, as well as to find and verify any new sites. The old sites from the home and SSRL datasets did not yield good electron density maps, so SOLVE was used to find new sites, which were verified against Patterson maps. New sites (from using solve_sad and solve_mad scripts) did not give traceable maps, nor did sites identified by CNS. The occupancy and B factor refinement continued to be a problem with SHARP. To address this, Uranium was used to model the heavy atom site, because it is the same MW as four iron atoms. This did not fix the problem. Molecular replacement in AmoRe and Phaser was tried using models derived from biotin synthase, but this was also unsuccessful. Finally, the alternate space group $P3_221$ was also tried (solve_sad and solve_mad) unsuccessfully.

Because the sites identified in the higher symmetry space groups were often at special positions, a lower symmetry space group was tried. The ALS data was reprocessed in $P3_1$ and very similar Patterson maps as those obtained in the higher symmetry space groups were calculated. In this space group, SOLVE identified two sites (see Table A2.5) that, upon refinement in SHARP, gave traceable maps. The secondary structural elements for the two molecules present in the $P3_1$ ASU were built, and an NCS matrix was eventually obtained for averaging. NCS averaging did not improve these maps, however. Phase combination with the secondary structural elements and phase extension with the native dataset also did not appear to improve the maps to a significant extent as judged by visual inspection.

| ALS | peak | inflection | remote | native1 |
|---|---|---|---|---|
| Beamline | 5.0.2 | 5.0.2 | 5.0.2 | 5.0.2 |
| Space Group | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ |
| Resolution (Å) | 2.6 | 2.75 | 2.75 | 2.7 |
| Rsym (%) | 7.8(30.3) | 7.6(26.9) | 7.5(25.6) | 7.2(28.2) |
| % refl rejected | 0.5 | 1.9 | 0.11 | 0.3 |
| Redundancy | 2.5 | 2.6 | 2.3 | 3.4 |
| I/sigma | 10.6(3.3) | 11.5(2.9) | 10.3(2.8) | 12.3(3.8) |
| % Complete | 97.4(89.0) | 98.4(94.3) | 93.9(85.4) | 90.9(88.8) |
| a (Å) | 57.814 | 57.922 | 57.975 | 57.533 |
| c (Å) | 116.528 | 116.651 | 116.651 | 115.719 |

Table A2.8: Data collection statistics in space group $P3_1$, ALS.

Crystals were then taken to California for the third time to collect more complete datasets.

## A2.II.B.5. Data from SSRL, $2^{nd}$ Trip

The crystals prepared for this trip were also grown from the conditions 0.1M tris pH 8.0, 26% PEG 3350 and cryoprotected with 20%MPD. Data were collected under the assumption that the space group was $P3_1$ to ensure a compete and highly redundant dataset.

| Dataset | $\lambda$ (A) | E(eV) | f' | f" | dist (mm) | exp time |
|---|---|---|---|---|---|---|
| Peak | 1.73542 | 7144.07 | -5.1913 | 4.0543 | 220 | 20 sec |
| inflection | 1.74166 | 7118.5 | -8.0985 | 1.9973 | 220 | 20 sec |
| Remote | 1.37755 | 9000 | -0.4 | 2.7 | 300 | 20 sec |
| Native1 | 0.97946 | 12387 | | | 335 | 30 sec |
| Native2 | 1.00002 | 12387 | | | 350 | 45 sec |

Table A2.9: Data collection information, SSRL2.

| Dataset | Peak | Inflection | Remote | Native1 | Native2 |
|---|---|---|---|---|---|
| Wavelength | 1.73542 | 1.74166 | 1.37755 | 1 | 1 |
| Beamline | 9.2 | 9.2 | 9.2 | 9.2 | 1.5 |
| Space Group | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ |
| Resolution (Å) | 2.87 | 3.2 | 3.7 | 2.25 | 2.4 |
| Rsym (%) | 6.8(30.2) | 6.5(27.9) | 7.8(27.4) | 5.6(36.8) | 4.1(21.3) |
| Total Obs | 107251(8059) | 77728(5886) | 47805(3609) | 83796(6334) | 76209(5813) |
| Unique Obs | 10145(750) | 7380(552) | 4805(356) | 20926(1577) | 16318(1571) |
| I/sigma | 29.5(4.7) | 30.7(5.3) | 25.2(7.0) | 19.2(2.5) | 30.4(5.6) |
| %Complete | 99.9(100) | 99.9(100) | 99.9(100) | 99.9(100) | 97.2(92.7) |
| Redundancy | 10.6(10.7) | 10.5(10.7) | 9.9(10.1) | 4.0(4.0) | 4.7(3.7) |
| %Comp(anom) | 99.3(98.8) | 99.9(100) | 99.7(100) | ---------- | ---------- |
| Redund (anom) | 4.9(4.9) | 5.3(5.3) | 4.8(4.7) . | ---------- | ---------- |
| a (Å) | 58.0858 | 58.2492 | 58.272 | 57.9772 | 57.005 |
| c (Å) | 117.2673 | 117.6435 | 117.8511 | 117.3712 | 116.662 |

Table A2.10: Data collection statistics in space group $P3_1$, SSRL2.

Because the new detector on beamline 9-2 at SSRL was not yet supported by HKL2000/Denzo/Scalepack, mosflm was used to process the data from that beamline.

The same sites (obtained using SAD data with SOLVE from the ALS peak) were refined in Sharp against the new data (both SAD and MAD data, cut off at 3.7Å resolution) to give good maps. The maps were two-fold NCS averaged and then used to further build a model of the two molecules. Later, all of the MAD data were included (to 2.87Å resolution) to refine the individual iron atoms of the cluster. This refinement yielded much better maps, into which most of the model (residues 5 to approximately 233) was eventually built.

Because the data scaled so well in $P3_121$ and $P3_221$, the space group was still in question. The data were reprocessed in $P3_121$, and molecular replacement was done into the SSRL2 native2 dataset with the hope that the cluster site coordinates for $P3_121$ could

be obtained from the MR solution. Though the site obtained this way matched the Patterson map, no good electron density maps were calculated with it. Next, the SSRL2 peak data was used to find the site in P3$_1$21 with SOLVE. This site did yield good electron density maps, which were also used to improve the model.

| Dataset | Peak | Inflection | Remote | Native1 | Native2 |
|---|---|---|---|---|---|
| Beamline | 9.2 | 9.2 | 9.2 | 9.2 | 1.5 |
| Space Group | P3$_1$21 | P3$_1$21 | P3$_1$21 | P3$_1$21 | P3$_1$21 |
| Resolution (Å) | 2.87 | 3.2 | 3.7 | 2.25 | 2.4 |
| Rsym (%) | 7.1(30.5) | 7.0(27.8) | 7.9(26.0) | 7.8(32.2) | 5.5(25.0) |
| I/sigma | 36.9(7.4) | 37.6(8.3) | 21.1(7.1) | 23.9(4.0) | 32.9(6.7) |
| %Complete | 100.0(100.0) | 99.9(100.0) | 99.8(100.0) | 99.0(93.7) | 98.2(95.9) |
| Redundancy | 10.8(10.9) | 10.7(10.8) | 5.0(4.9) | 9.6(4.4) | 3.5(6.8) |

Table A2.11: Data collection statistics in space group P3$_1$21, SSRL2.

## A2.II.C. Refinement of the substrate-free PflAE model

At this point, most of the residues and approximately 40% of the sidechains had been incorporated into the current model, with the exception of the loop following β6, which had been built with only 10-residues. Parameters for the cluster were obtained and verified in both CNS and SHELXL formats. Refinement was carried out in P3$_1$21 and P3$_1$ in CNS, but the R factors either remained high or indicated overfitting of the data. For example, refinement of the model against one of the native datasets in space group P3$_1$21 resulted in an R and Rfree stuck at approximately 37 and 46. Refining in both space groups in CNS has not brought the R factors down.

Several factors can cause this type of problem. High R factors such as these can result from twinning, refinement in an incorrect space group, poor quality data, or other problems. Each of these possibilities was investigated.

### A2.II.C.1. Twinning

Twinning was considered as a potential problem. Both the Yeates and CNS twinning tests showed that the distribution of intensities did not suggest twinning, but gave results that suggested almost perfect merohedral twinning (twin operator h, -h-k, -l), though that specific twin operator would generate P3$_1$21 symmetry, and identification of this twin operator was probably an artifact of the higher symmetry (meaning, the crystals were actually P3$_1$21). The same results were obtained with other data available in our lab (BioB and thioredoxin data) that have been used to successfully determine crystal structures in P3$_1$21, suggesting that the apparent twinning was an artifact of higher crystallographic symmetry and that twinning was not the real problem. Refinement in CNS of two molecules in P3$_1$ results in R=23 and Rfree=42, but refinement in SHELXL in P3$_1$ with the following commands:

TWIN 1 0 0 -1 -1 0 0 0 -1
BASF 0.5 (approximate value – refined by program)

eventually gave much better R factors (R=28, Rfree=34). Though the R factors were greatly improved with this refinement technique in SHELXL, it was still unclear that twinning was the cause of our high R factors.

The data were detwinned using CCP4 (keeping the anomalous signal by detwinning I+ and I- separately) and used to generate maps, find the sites, and refine the model. The anomalous signal was much lower after detwinning. Maps from SOLVE were untraceable, and the two sites were closer than expected in comparison to the sets of sites found previously. Maps generated by SHARP with old sites were much noisier than what had been used before, but did contain some secondary structural elements. Refinement against detwinned data also resulted in high R factors. Introduction of more noise and reduction in the anomalous signal was expected, because the twin fraction used was so high. Again, these experiments did not convince us that these data were twinned.

*A2.II.C.2. Data reprocessing in order to improve dataset quality*

Several strategies were pursued to determine the source of the refinement difficulties unequivocally. First, because poor data quality could contribute to the high R factors, the data were reprocessed to improve statistics such as $\chi^2$ and $R_{sym}$. Second, self-rotation functions were calculated for data processed in space groups C2 and P31 in an effort to distinguish between the three space group possibilities. Next, all of the SSRL2 data were reprocessed in C2 and used to obtain new experimental maps in order to see whether use of the lower symmetry space group would change the maps in any way. Finally, the experimental model was refined against native data processed in various other space group possibilities.

In order to address the possibility that the problematic refinement was caused by poor quality data alone, we reprocessed SSRL2 data with HKL2000 in order to work with cleaner datasets. Following reprocessing in space group $P3_1$ and $P3_121$, the R factors remained high, and strangely, maps (experimental and model-biased) calculated from these datasets (using old sites) were not as good as maps calculated from the mosflm-processed datasets.

The data were then reprocessed in multiple space groups. A self-rotation function was carried out with molrep in both space groups C2 and $P3_1$. The output indicated the presence of a crystallographic (or very close to crystallographic) 3-fold and three crystallographic (or, again, very close to crystallographic) 2-folds, as would be expected for space group $P3_121$ or $P3_221$. These results again suggested that the correct space group is $P3_121$ or $P3_221$. Still, the models were refined in each space group to determine the correct space group unequivocally.

*A2.II.C.2.a. C2*

The first SSRL2 native dataset (n1) was processed in space group C2 (see Table A2.12 for statistics) and used for a molecular replacement search. A solution with reasonable packing and orientations reminiscent of those observed in the trigonal space groups was obtained using molrep. The three MAD datasets were reprocessed in C2 with

218

HKL2000 (see Table A2.12), and consistent indexing was verified, though merging R factors indicated poor agreement between datasets, particularly the remote wavelength data (Table A2.13).

| Dataset | Peak | Inflection | Remote | Native1 |
|---|---|---|---|---|
| Beamline | 9.2 | 9.2 | 9.2 | 9.2 |
| Space Group | C2 | C2 | C2 | C2 |
| Resolution (Å) | 2.87 | 3.2 | 3.7 | 2.25 |
| Rsym (%) | 7.2(32.7) | 7.0(32.5) | 8.4(31.9) | 3.9(33.6) |
| I/sigma | 13.3 (3.1) | 15.7(4.1) | 12.8(3.9) | 19.8(2.8) |
| %Complete | 98.4(97.6) | 99.3(99.9) | 98.2(100.0) | 96.1(94.2) |
| Redundancy | 3.6 | 3.7 | 3.6 | 2.7 |
| a (Å) | 100.649 | 100.844 | 100.807 | 100.438 |
| b (Å) | 58.089 | 58.293 | 58.445 | 58.073 |
| c (Å) | 117.306 | 117.565 | 117.623 | 117.5 |
| beta (°) | 90.011 | 89.961 | 89.874 | 90.034 |

Table A2.12: Data collection statistics in space group C2, SSRL2.

| | Peak | Inflection | Remote | Native1 |
|---|---|---|---|---|
| Peak | | 11.2 | 22.3 | 8.9 |
| Inflection | 11.2 | | 14.8 | 15.7 |
| Remote | 22.3 | 14.8 | | 24.4 |
| Native1 | 8.9 | 15.7 | 22.4 | |

Table A2.13: Merging Rfactors (%) for MAD data in C2, SSRL2.

The three iron sites obtained from the molecular replacement solution (estimated by moving an atom into the center of the cluster's density) were used to generate experimental maps with Sharp. There was a problem with site 2, which did not refine properly in Sharp. Experimental maps were also generated from peaks identified by SOLVE (using SAD with the peak data, Table A2.12, A2.13). The two sets of three sites were later confirmed as identical.

| SITES | x | y | z |
|---|---|---|---|
| MR | 0.118 | 0.782 | 0.824 |
| | 0.798 | 0.973 | 0.492 |
| | 0.921 | 0.387 | 0.843 |
| | | | |
| SOLVE | 0.883 | 0.499 | 0.175 |
| | 0.082 | 0.101 | 0.158 |
| | 0.299 | 0.249 | 0.491 |

Table A2.14: Sites obtained for SAD data in C2, SSRL2.

The phasing statistics presented in Table A2.15 were obtained by refinement of the two sets of iron sites in Sharp to 3.7Å resolution. Both sets of sites gave traceable experimental maps, but the quality of those maps was not as high as those obtained in the trigonal space groups. Three molecules were fit into both maps (molecules were fit into SOLVE maps by hand, and the MR solutions fit the maps calculated with those sites –

the maps were not superimposable) and the three-fold screw axis was identified in both sets of solutions.

| SITES | x | y | z | FOM | PP | occ problems? |
|---|---|---|---|---|---|---|
| MR | 0.118 | 0.782 | 0.824 | | | |
| | 0.798 | 0.973 | 0.492 | 0.62 | 1-1.15 | No |
| | 0.921 | 0.387 | 0.843 | | | |
| | | | | | | |
| SOLVE | 0.883 | 0.499 | 0.175 | | | |
| | 0.082 | 0.101 | 0.158 | 0.64 | 1.0-1.381 | No |
| | 0.299 | 0.249 | 0.491 | | | |

Table A2.15: Phasing statistics to 3.7 Å resolution for the sites obtained from MAD data in C2, SSRL2.

Effort was put into improving experimental maps by improving the quality of the remote dataset. Suspicious frames were removed, scale and B factors were restrained or allowed to refine, and fitting of parameters was done by frame or by batch. The remote data's agreement with the other 3 datasets did not improve, and similarly the experimental maps did not improve significantly.

A new test set was created for this dataset and the model was refined in C2 with SHELXL to 2.25 Å resolution, giving R=36.67, Rfree=47.05.

*A2.II.C.2.b. P1*

The first SSRL2 native dataset (Table A2.11) was reprocessed again, this time in P1 with HKL2000. The statistics were as follows: Rsym = 3.1(30.1) % to 2.25 Å, I/sigma = 17.3(2.4), redundancy = 1.6 and completeness = 80.6(78.3)%. A solvent content of 39.1% meant the cell would contain 6 molecules per asymmetric unit, and the cell angles and lengths (57.999 x 58.071 x 117.506, 89.994, 90.034, 120.19) were almost identical to the trigonal cell. Molecular replacement with AmoRe and molrep was unsuccessful. To place the molecules appropriately in the cell, the six symmetry related molecules within the P3$_1$21 cell were written into pdb files. The six molecules were refined in CNS (no ncs restraints) with best Rfactors being R=46.53, Rfree=54.49.

*A2.II.C.2.c. R3 / R3$_2$*

Some effort was put into reprocessing the SSRL2 data in the rhombohedral space groups. The first SSRL2 native dataset Table A2.11) was again used to test these space groups. The distortion index given after indexing several images from the SSRL2 native1 and peak (SSRL2, Table A2.11) datasets gave a higher value (~6%) for the rhombohedral lattice than acceptable. The SSRL2 native1 dataset (Table A2.11) was processed using this lattice anyway, and the Rsym values obtained from multiple attempts at processing in R3 was also unacceptably high (~20% overall). Molecular replacement was attempted briefly with the native1 data integrated in P3 and scaled in R3, but those attempts were unsuccessful. According to the distortion index for this lattice, these space groups are very unlikely.

**A2.III. The substrate-bound crystal form**

Luckily, a new crystal form was identified using a sample containing the fully reconsituted enzyme, AdoMet, and a seven-residue peptide substrate (RVSGYAV). The optimized crystals were typically small but diffracted to (at best) 2.7 Å at the synchrotron.

## A2.III.A. Crystallization of substrate-bound PflAE

Upon having difficulty during refinement of the protein model against data from the "trigonal" crystal form, several new protein preps, with different substrates added, were used to screen for new anaerobic crystallization conditions. Six different stocks were received from Jian Yang in Joan Broderick's lab:

1. Activase (AE) alone
2. AE with 10x AdoMet
3. AE with 10x AdoMet and peptide
4. AE with 10x AdoMet and YfiD
5. AE with 10x AdoMet and PFL
6. AE alone in a different buffer



Figure A2.7: Substrate-bound PflAE crystal form.
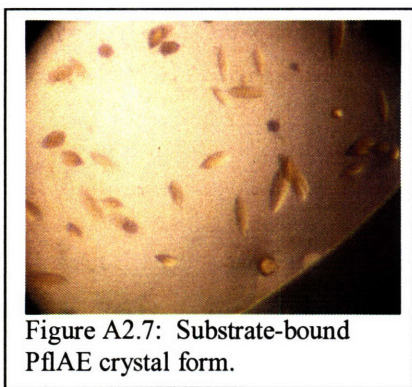
Sparse matrix screens (from Hampton Research) were used as a starting point to obtain a new crystal form. Focus was placed on samples 2, 3 and 4 first, eventually yielding four new possible hits:

1. Sample 2: HIS#45: 0.1M tris pH 8.5, 25% PEG 3350 – old conditions
2. Sample 2: CS1#2 over 2.0$\underline{M}$ $(NH_4)_2SO_4$: 0.4M K, Na tartrate
3. Sample 2: CS1#37 over 2.0$\underline{M}$ $(NH_4)_2SO_4$: 8% PEG 4K, 0.1M NaOAc pH 4.6
4. Sample 3: HIS#25: 3.5$\underline{M}$ sodium formate pH 7.0
5. Sample 4: CS1#43 over 2.0$\underline{M}$ $(NH_4)_2SO_4$: 30% PEG 1500

The crystals obtained in HIS#25 with Sample 3 (the fourth hit listed above) were reproduced and optimized to yield diffracting crystals. The components that were optimized included buffer identity and pH, salt identity and concentration, drop volume, drop ratio, well volume, well identity, various additives, and addition of microseeds. Small crystals formed reproducibly, and the main hurdle to overcome to generate diffracting crystals was their small size. The techniques used to increase crystal size

were addition of detergents to both hanging and sitting drops, microseeding, and a combination of microseeding and detergents. Large crystals could not be reproduced consistently – apparently the factor influencing crystal size was independent of the detergent and presence of seeds, though these techniques did seem to help. The largest crystals (the ones used to collect data at SSRL) were all harvested from the same drop (Tray label: AE+pept Reprod2, well number C4, both hanging and sitting drop, set up on 4/7/06).

The final crystallization conditions used to generate the crystals used for data collection were as follows. Drops were set up either sitting on microbridges or hanging on a coverslip against 500mL well solution consisting of 0.1$\underline{M}$ hepes pH 6.8, 3.5$\underline{M}$ sodium formate. The drops contained 1μL protein stock solution (with protein at 20 mg/mL), 0.8μL well solution, and 0.2μL methyl-4-heptyl-β-D-maltoside (Hampton Detergent Screen 2 #3). Crystals took approximately six days to appear.

## A2.III.B. SSRL3 data collection and phasing of substrate-bound PflAE
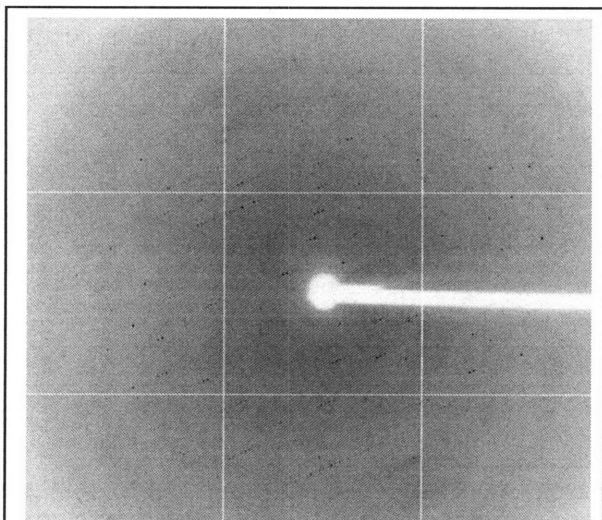
*A2.III.B.1. Data collection*



Figure A2.8: Representative diffraction image of substrate-bound PflAE from SSRL beamline 9-2.

Three MAD datasets and five native datasets were collected at SSRL and processed in space group P6$_1$22.

| Dataset | λ (A) | E(eV) | f' | f" | dist (mm) | exp time |
|---|---|---|---|---|---|---|
| Peak1 | 1.73948 | 7128.170 | -6.456790 | 4.416056 | 210 | 25 |
| Rem1 | 1.36241 | 9100.000 | -0.300000 | 2.600000 | 300 | 25 |
| Peak2 | 1.73950 | 7127.310 | -6.412502 | 4.222880 | 220 | 25 |
| Infl2 | 1.74141 | 7119.530 | -7.824151 | 2.466863 | 220 | 25 |
| Rem2 | 1.36241 | 9100.000 | -0.300000 | 2.600000 | 300 | 25 |
| Peak3 | 1.73955 | 7127.140 | -6.410355 | 4.212681 | 210 | 20 |
| Infl3 | 1.74158 | 7118.840 | -7.940308 | 2.435220 | 210 | 25 |
| Rem3 | 1.36241 | 9100.000 | -0.300000 | 2.600000 | 300 | 15 |
| Native1 | 1.00000 | 12387 | ————— | ————— | 300 | 20 |
| Native2 | 1.00000 | 12387 | ————— | ————— | 300 | 35 |
| Native3 | 0.95000 | 13050.480 | ————— | ————— | 300 | 30 |
| Native4 | 0.95000 | 13050.480 | ————— | ————— | 450 | 45 |
| Native5 | 0.95000 | 13050.480 | ————— | ————— | 450 | 30 |

Table A2.16: Data collection information from substrate-bound PflAE crystals, SSRL3.

MAD data: Peak1 and Remote1 were collected on the same crystal, and suffered decay. Peak2, Infl2 and Rem2 were each collected on the same crystal; 50° of each wavelength was collected, but decay was again a problem. Finally, Peak3 (PXG2), Infl3 (PXG3) and Rem3 (PXH1) were each collected on a separate crystal to minimize problems from decay. This third MAD dataset was used to generate experimental maps.

The datasets used to generate experimental maps and to later refine the model were Peak3, Infl3, Rem3 and Native2 Table A2.16, A2.17). The data processing statistics for each of those datasets are compiled in Table A2.17.

| Dataset | Peak3 | Inflection3 | Remote3 | Native2 |
|---|---|---|---|---|
| Wavelength | 1.73955 | 1.74158 | 1.36241 | 1.000 |
| Beamline | 9-2 | 9-2 | 9-2 | 9-2 |
| Space Group | P6$_1$22 | P6$_1$22 | P6$_1$22 | P6$_1$22 |
| Resolution (Å) | 2.9 | 2.8 | 2.7 | 2.8 |
| Rsym (%) | 9.8(42.0) | 8.1(36.9) | 7.3(35.9) | 5.2(35.6) |
| Total # Refl | 381294 | 145346 | 154570 | 101174 |
| # rej | 825 | 646 | 461 | 59 |
| % rej | 0.22 | 0.44 | 0.30 | 0.06 |
| Unique Obs | 12853(1152) | 12677(1107) | 14585(1406) | 8085(731) |
| I/sigma | 18.6(2.7) | 13.1(3.1) | 19.9(3.1) | 25.4(3.63) |
| %Complete | 98.5(87.6) | 88.7(78.6) | 98.6(95.8) | 98.7(94.3) |
| Redundancy | 9.8(5.8) | 5.4(5.0) | 5.5(4.0) | 5.5(4.4) |
| a (Å) | 74.56 | 74.408 | 74.426 | 74.35 |
| c (Å) | 187.979 | 187.456 | 187.608 | 187.453 |

Table A2.17: Data processing statistics of SSRL3 – substrate-bound PflAE.

The data collected from this P622 crystal form, while lower resolution than hoped for, were of much higher quality than any of the data from crystal form 1. The data indexes in P622 easily, has reasonable mosaicities (Table A2.18), and scales together with reasonable R factors.

| Dataset | Crystal | Beam | λ | Res (Å) | Rsym (%) | Complete | Redund | I/sigma | Mos (Å) |
|---------|---------|------|------|------|-----------|-----------|-----------|-----------|---------|
| Native1 | PXI2 | 9-2 | 1 | 2.75 | 8.1(38.0) | 96.5(77.0) | 9.9(5.7) | 22.2(2.39) | 0.4-0.8 |
| Native2 | PXB1 | 9-2 | 1 | 2.8 | 5.2(35.6) | 98.7(94.3) | 5.5(4.4) | 25.4(3.63) | 0.3 |
| Native3 | H8 | 9-1 | 0.95 | 3.3 | 10.9(31.7) | 99.6(99.0) | 13.5(11.0) | 24.3(6.0) | 0.5 |
| Native4 | PXH1 | 9-1 | 0.95 | 3.3 | 10.4(40.7) | 99.6(100) | 10.1(10.1) | 20.8(5.9) | 0.4-0.8 |
| Native5 | PXH2 | 9-1 | 0.95 | 3 | 7.9(43.9) | 99.9(99.8) | 19.6(17.3) | 40.5(6.4) | 0.2-0.3 |

Table A2.18: Data processing statistics for each SSRL3 native dataset – substrate-bound PflAE.

## A2.III.B.2. Obtaining phases for substrate-bound PflAE

Both molecular replacement and MAD techniques were used to attempt to phase the data. One site corresponding to the iron-sulfur center was identified by SOLVE in space group $P6_122$ with the MAD3 data, and refined in Sharp to generate experimental maps to 2.7 Å resolution. The model built and partially refined against the first crystal form was fit manually into the new experimental maps and rebuilt to match the new data.

### A2.III.B.2a. Phasing the data by MR

Six different molecular replacement search models derived from the crystal form 1 model were used to search for a solution from the native2 dataset. The six models (1 – full chain, no cluster; 2 – full chain, with cluster; 3 – no loop6b, no cluster; 4 – no loop 6b, with cluster; 5 – just barrel [20-157 only]; 6 – poly ala, no loop) each gave one of two solutions with good Z scores. The two different solutions obtained were related by a translation along c, due to the ambiguity in assigning z in trigonal space groups. However, the molecular replacement solutions yielded high and overfit R factors when refined in CNS. Because of high R factors and poor 2fo-fc density maps, I turned to the MAD3 data to phase the substrate-bound PflAE data.

### A2.III.B.1b. Phasing with the MAD3 data

Peak3 data yielded good anomalous difference Fourier maps with little noise and strong signal. SOLVE identified one site (0.656, 0.194, 0.051) corresponding to the iron sulfur cluster. The data were scaled (see Table A2.19) in CCP4 with reasonable R factors between datasets, considering that the data came from three separate crystals.

| To 3.5 Å resolution | Peak | Infl | Rem |
|---------------------|------|-------|-------|
| Peak | ------- | 0.167 | 0.135 |
| Infl | 0.167 | ------- | 0.099 |
| Rem | 0.135 | 0.099 | ------- |

| All data | Peak | Infl | Rem |
|----------|------|-------|-------|
| Peak | ------- | 0.182 | 0.135 |
| Infl | 0.182 | ------- | 0.11 |
| Rem | 0.135 | 0.11 | ------- |

Table A2.19: Scaling statistics (%) for the MAD data, SSRL3.

Sharp refined the site to 3.5 Å resolution and solvent flattened the density using an optimized solvent content of 48.6%. After some fitting of the model and refinement (which gave R factors in the upper 40s), the individual iron sites were placed and refined to 2.77 Å resolution.

Phasing Power

| | FOM(acen/cen) | $P_{Anom}$ | $I_{Anom}$ | $R_{Anom}$ | $P_{Iso}$ | $I_{Iso}$ |
|---|---|---|---|---|---|---|
| To 3.5 Å | 0.53774/0.23406 | 1.638 | 1.264 | 1.622 | 0.423 | 0.317 |
| To 2.77 Å | 0.45737/0.29772 | 2.049 | 1.135 | 1.659 | 0.429 | 0.918 |

Table A2.20: Phasing statistics yielded from SSRL3 MAD3 data.

## A2.III.C. Refinement of substrate-bound PflAE

The native2 dataset was used to refine the model to reasonable Rfactors. The current model has decent geometry (Table A2.21).

The model obtained from the first crystal form was dragged into the new experimental maps and modified to match the new electron density. The major adjustments involved were the loop preceding strand $\beta 1$ and the loop following strand $\beta 6$. Rounds of refinement in CNS and rebuilding in Xfit were carried out to R factors of 32.45/36.11. Because initially refinement against native2 resulted in high, overfit R factors, the first rounds of refinement in CNS were done against the MAD3 data to 3.0 Å with Hendrickson-Lattman phases included. Phase combined maps calculated in CNS were used in conjunction with the experimental maps to rebuild the model. The peptide was added after all of the ordered sidechains were fit into density, and refinement of the protein model with the peptide was carried out against the MAD data in order to properly orient the peptide.

The peptide direction was analyzed to confirm its orientation. Points considered during this analysis included careful examination of the electron density, R factors, and analysis of the resulting model in terms of reasonable hydrogen bonding interactions, an orientation that observed in Pfl, consistency with results known from biochemical experiments on this system, and the accessibility of the pro-S hydrogen atom of $G_{734}$ to the C5' atom of AdoMet.

## A2.III.D. Reanalysis of substrate-free PflAE

The new model (refined against the substrate-bound data) was rebuilt and refined against the apo data, yielding better R factors and demonstrating that errors in the model due to poor experimental maps and data quality were responsible for the high R factors observed previously, and not twinning.

### A2.III.D.1. Initial refinement rounds with unadjusted model

Different approaches to refinement were investigated with the substrate-bound PflAE protein model, using only the protein chain and the 4Fe-4S cluster. The substrate-free PflAE native1 dataset from SSRL2 was used, and the data was not detwinned or in any

way manipulated. The different approaches used and the resulting R factors are summarized here:

| | |
|---|---|
| Shelxl P3$_1$ modeled as twinned | 22 / 27 |
| Shelxl P3$_1$ not modeled as twinned | 34 / 42 |
| Shelxl P3$_1$21 not twinned | 59 / 56 |
| CNS P3$_1$ not twinned | 30 / 41 |
| CNS P3$_1$ twinned | 23 / 34 |
| CNS P3$_1$21 not twinned | 46 / 54 |

However, no twin tests confirmed twinning, including SCALA in CCP4, the CNS twin test, the Yeates Twinning Server, and the Crystal Twinning Server based on Yeates and Padilla's H-test.

## A2.III.D.2. Refinement of a new substrate-free PflAE model: Comparison of different datasets / space groups

All five datasets from PflAE crystal form 1 were reprocessed and scaled in the following spacegroups: C2 (Table A2:22), P3$_1$ (Table A2:23) and P3$_1$21 (Table A2:24). The new substrate-free PflAE model was used as a search model for molecular replacement into each dataset scaled in each spacegroup (15 in total, all using the same search model). Each solution was then refined to 2.8 Å resolution in CNS against each reflection file in the following way: Rigid / Minimize / Bindividual / Anneal / Bindividual. NCS restraints were used when possible. The data:parameter ratios were determined taking NCS restraints (segid A = segid B for P3$_1$ and segid A = segid B = segid C for C2, nothing for P3$_1$21) into account.

### C2

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| R | ———— | ———— | ———— | 25.80 | 24.12 |
| Rfree | ———— | ———— | ———— | 27.72 | 28.94 |
| data:param | ———— | ———— | ———— | 2.19 | 2.09 |

### P3$_1$

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| R | 28.50 | 28.99 | 29.65 | 26.03 | 26.21 |
| Rfree | 34.48 | 31.79 | 33.60 | 29.66 | 30.81 |
| data:param | 2.66 | 2.86 | 2.88 | 2.90 | 2.84 |

### P3$_1$21

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| R | 28.70 | 30.52 | 31.30 | 25.58 | 26.33 |
| Rfree | 37.01 | 36.22 | 35.59 | 32.28 | 35.42 |
| data:param | 1.40 | 1.48 | 1.48 | 1.46 | 1.44 |

These results demonstrate that the true space group of the substrate-free PflAE data is P3$_1$21. In addition to this, the reprocessing of each dataset likely improved the data quality significantly. Further rounds of refinement were carried out using dataset n1_ssrl2 in space group P3$_1$21 to reasonable R factors and geometry (Table A2.21).

226

|  | AE | Pept-AE |
|---|---|---|
| Resolution Limits (Å) | 38.19 - 2.25 | 29.13 - 2.77 |
| # Unique Reflections | 19688 | 8376 |
| # Reflections - Test Set | 1869 | 438 |
| $R_{work}$ (%) | 23.9 | 22.9 |
| $R_{free}$ (%) | 32.3 | 26.1 |
| Average B factor ($\text{Å}^2$) | 56.3 | 80 |
|   Protein | 55.7(1902) | 76.9(1872) |
|   4Fe-4S | 44.3(8) | 59.7(8) |
|   AdoMet | N/A | 75.7(27) |
|   Peptide | N/A | 104.4(42) |
| # protein atoms | 1902 | 1914 |
| # water atoms | 49 | 16 |
| # non-protein atoms | 8 | 50 |
| Bond length deviation (Å) | 0.006 | 0.008 |
| Bond angle deviation (°) | 1.4 | 1.5 |
| Ramachandran Plot | | |
|   Residues in allowed regions (%) | 84.6 | 84.3 |
|   Res in generously allowed region (%) | 14.5 | 15.3 |
|   Res in additional allowed region (%) | 0.5 | 0.5 |
|   Res in disallowed region (%) | 0.5 | 0 |

Table A2.21: Refinement statistics, both substrate-free (AE) and substrate-bound (pept-AE) forms. Numbers in parenthesis refer to number of atoms.

Table A2:22. Data statistics — natives from substrate-free PflAE crystals — processed in C2. Numbers in brackets refer to the low resolution bin reflections, and numbers in parenthesis refer to the high resolution bin reflections. Bad rejection pattern refers to whether a strange pattern is observed in the rejections, which arises from poor agreement between symmetry-related reflections. Rej remaining refers to any rejections still listed in the scalepack output file after multiple rounds of scaling.

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| SG | C2 | C2 | C2 | C2 | C2 |
| Resolution (A) | 50-2.7 | 50-2.65 | 50-2.8 | 50-2.25 | 50-2.5 |
| Rsym (%) | [18.0]21.8(61.7) | [8.1]12.1(42.5) | [5.4]13.0(45.4) | [2.7]3.7(33.2) | [3.1]4.6(14.1) |
| Total Obs | | 32875(3098) | 57261(3981) | 80121(7379) | 85093(5255) |
| Unique Obs | | 17118(1696) | 15640(1644) | 30887(3060) | 20774(1868) |
| I/sigma | | 8.156(2.890) | 12.7(3.72) | 18.1(2.52) | 25.50(6.025) |
| % complete | | 85.5(84.9) | 92.5(97.5) | 95.5(94.0) | 89.9(82.9) |
| Redundancy | | 1.9205(1.827) | 3.66(3.63) | 2.6(2.4) | 4.10(2.80) |
| a (A) | 59.073 | 58.441 | 58.566 | 100.425 | 98.724 |
| b (A) | 100.5 | 101.196 | 101.423 | 58.065 | 56.993 |
| c (A) | 117.034 | 117.338 | 117.455 | 117.484 | 116.602 |
| β (°) | 90.811 | 89.985 | 89.997 | 90.036 | 89.956 |
| # refl rejected | 19290 | 11108 | 31818 | 974 | 6773 |
| total refl | 131576 | 222096 | 542109 | 398774 | 767783 |
| % refl rejected | 14.70 | 5.00 | 5.87 | 0.2 | 0.9 |
| mosaicty (°) | | 1.4-2.0 | 1.2-1.6 | ~1 | ~1.8 |
| chi^2 | too high | 3.33 | 2.26 | 1.00 | 0.98 |
| error scale | 2.4+ | too high! | bad chi2 by res | 0.9 | 0.95 |
| scaling b factors | | -3 to 2 | | grad incr | 0-6, some high |
| scale factors | | 0.8 - 0.4 - 0.8 | | 0.9 steady | ~1, some high |
| syst abs present | | no | no | no | no |
| bad rej pattern | | low red | ? | no - low red | |
| rej not thrown | | lots - not done | lots | none | 0 |

Table A2:23. Data statistics — natives from substrate-free PflAE crystals — processed in $P3_1$. Numbers in brackets refer to the low resolution bin reflections, and numbers in parenthesis refer to the high resolution bin reflections. Bad rejection pattern refers to whether a strange pattern is observed in the rejections, which arises from poor agreement between symmetry-related reflections. Rej remaining refers to any rejections still listed in the scalepack output file after multiple rounds of scaling.

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| SG | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ |
| Resolution (A) | 50-2.7 | 50-2.65 | 50-2.8 | 50-2.25 | 50-2.4 |
| Rsym (%) | [6.0]7.1(28.2) | [3.8]7.5(31.3) | [4.5]9.5(37.5) | [4.8]6.4(40.6) | [3.1]4.1(21.3) |
| Total Obs | 36610(3178) | 34737(2822) | 54787(4116) | 82654(7502) | 76209(5820) |
| Unique Obs | 10769(1020) | 12355(1184) | 1044(1103) | 20500(2065) | 16318(1571) |
| I/sigma | 12.35(3.76) | 12.2(3.6) | 15.01(4.25) | 14.6(3.39) | 30.4(5.6) |
| % complete | 90.9(88.8) | 94.5(92.4) | 94.1(98.5) | 98.1(97.8) | 97.2(92.7) |
| Redundancy | 3.4(3.1) | 2.9(2.4) | 5.2(3.73) | 4.0(3.6) | 4.7(3.7) |
| a (A) | 57.533 | 58.432 | 58.511 | 58.052 | 57.055 |
| b (A) | 57.533 | 58.432 | 58.511 | 58.052 | 57.055 |
| c (A) | 115.719 | 117.358 | 117.432 | 117.464 | 116.662 |
| $\beta$ (°) | ------ | ----- | ------ | ---- | ----- |
| # refl rejected | 897 | 1831 | 2872 | 1089 | 2596 |
| total refl | 273730 | 220827 | 463485 | 396591 | 406725 |
| % refl rejected | 0.3 | 0.8 | 0.6 | 0.27 | 0.6 |
| mosaicity (°) | 1.5-2.0 | 1.4-2.0 | 1.1-1.6 | ~1 | ~1.8 |
| chi^2 | 1.044 | 1.052 | 1.286 | 0.991 | 1.08 |
| error scale | 1.4 | 1.4 | 1.4 | 1.5 | 0.8 |
| bfactors | vary widely | good | good | 0-12 | range 0-3 |
| scale factors | good | drop+restore | good | slow decline | incr slowly 1-1.6 |
| syst abs present | yes | yes | yes | yes | yes |
| bad rej pattern | few - low red | yes? Low red | yes | yes | sortof |
| rej not thrown | none | 3 | ~30 | 4 | 5 |

Table A2:24. Data statistics – natives from substrate-free PIIAE crystals – processed in P3₁21. Numbers in brackets refer to the low resolution bin reflections, and numbers in parenthesis refer to the high resolution bin reflections. Bad rejection pattern refers to whether a strange pattern is observed in the rejections, which arises from poor agreement between symmetry-related reflections. Rej remaining refers to any rejections still listed in the scalepack output file after multiple rounds of scaling.

| | n1_als | n1_ssrl1 | n2_ssrl1 | n1_ssrl2 | n2_ssrl2 |
|---|---|---|---|---|---|
| SG | P3₁21 | P3₁21 | P3₁21 | P3₁21 | P3₁21 |
| Resolution (A) | 50-2.7 | 50-2.65 | 50-2.8 | 50-2.25 | 50-2.4 |
| Rsym (%) | [6.5]7.9(31.2) | [4.0]7.5(35.2) | [4.7]10.3(39.6) | [4.9]6.6(43.3) | [3.3]5.5(24.9) |
| Total Obs | 36763 | 34676 | 54774 | 82763 | 75621 |
| Unique Obs | 6252(589) | 6963(669) | 5896(591) | 11183(1103) | 9015(861) |
| I/sigma | 153.(5.0) | 16.3(4.1) | 16.93(5.667) | 18.7(4.6) | 39.67(7.30) |
| % complete | 95.4(95.2) | 96.6(97.1) | 95.9(99.0) | 98.2(98.7) | 98.2(95.9) |
| Redundancy | 5.9 | 5.0 | 9.29 | 7.4 | 8.388 |
| a (A) | 57.548 | 58.441 | 58.512 | 58.051 | 57.055 |
| b (A) | 57.548 | 58.441 | 58.512 | 58.051 | 57.055 |
| c (A) | 115.726 | 117.366 | 117.436 | 117.466 | 116.664 |
| b (°) | ---- | ----- | ------ | ---- | ----- |
| # refl rejected | 895 | 2374 | 2931 | 1163 | 2766 |
| total refl | 273730 | 220827 | 463485 | 396591 | 406725 |
| % refl rejected | 0.3 | 1.1 | 0.811 | 0.29 | 1.1 |
| mosaicity (°) | 1.5-2.0+ | 1.4-1.9 | 1.1-1.6 | ~1 | ~1.8 |
| chi^2 | 1.061 | 1.031 | 1.166 | 0.956 | 1.179 |
| error scale | 1.4 | 1.4 | 1.4 high by res | 1.5 | 0.8 (by res high) |
| bfactors | -5 to 3, outliers | smooth, 0-4 | good | grad incr 0-11 | -1 to 5 |
| scale factors | good | drop+restore | 1.0 - 0.8 - 1.0 | slow decline | all ~1 |
| syst abs present | yes | yes | yes | yes | yes |
| bad rej pattern | some | some | ? | yes | yes |
| rej not thrown | 1 | 0 | ~30 | 1 | 1 |

230

# Jessica L. Vey

143 Albany St. Apt 401A      857-928-4256

Cambridge, MA 02139      jessvey@mit.edu

## EDUCATION

- Massachusetts Institute of Technology (MIT), Ph.D. in Biological Chemistry, February, 2007.
- Temple University, B.S. *magna cum laude* in Chemistry, May 2001.

## RESEARCH EXPERIENCE

- **Graduate Research Assistant** October, 2001 – present (MIT)

My thesis work with Prof. Catherine Drennan focused on solution of X-ray crystal structures of pyruvate-formate lyase activating enzyme (PflAE) in the substrate-free and substrate-bound forms. PflAE is an AdoMet radical enzyme responsible for the activation of pyruvate-formate lyase (PFL), a glycyl radical containing enzyme (GRE) central to bacterial anaerobic glucose metabolism.

- **Undergraduate Research Assistant,** 2000 – 2001 (Temple University)

My research with Prof. Trevor Douglas aimed to purify and characterize the viral coat protein of Cowpea Chlorotic Mottle Virus (CCMV) as a potential drug delivery vehicle.

- **Undergraduate Research Assistant,** 1999-2000 (Temple University)

In the laboratory of Prof. David Dalton I carried out the initial steps of the synthetic route to the four stereoisomeric 3-hydroxy-2-hydroxymethylpyrrolidines. This study demonstrated a synthetic route to "aza-DNA" compounds.

## AWARDS

William Asbjornsen Albert Memorial Fellowship, MIT, 2003-2005
Corning Fellowship, MIT, 2002-2003
MIT Presidential Fellowship, 2001-2002
Temple University Outstanding Academic Student Scholarship, 1997-2001
Melvin P. Klein Professional Development Award, 2007
Morse Travel Grant, 2007
Ludo Frevel Crystallography Scholarship Award, 2006
Morse Travel Grant, 2006
Eli Lilly Travel Grant, 2003
ACS Achievement Award, Temple University, 2001
Merck Index Award, Temple University, 2001
Cotton and Tulinsky Prize, Temple University, 2001
Phi Beta Kappa Honor Society, Member, 2001
ACS Award in Undergraduate Analytical Chemistry, TU, 2000
Golden Key National Honor Society, Member, 1999
Temple University Honors College Degree , 1998

## PUBLICATIONS

Vey J, Drennan CL. Structural Insights into Radical Generation by the AdoMet Radical Superfamily. *(Manuscript in Preparation)*

Vey J, Yang J, Li M, Broderick JB, Drennan CL. Structural Basis for Glycyl Radical Formation by Pyruvate Formate-lyase Activating Enzyme. *(Manuscript In Preparation)*

Tirupati B, Vey J, Drennan CL, Bollinger JM. Kinetic and Structural Characterization of Slr0077/SufS, the Essential Cysteine Desulfurase from Synechocystis sp. PCC 6803. (2004) *Biochemistry.* 43(38): 12210 – 12219.

Mascavage LM, Lu Q, Vey J, Dalton DR, Carroll JR. Enantioselective Synthesis of Aza Sugars from Amino Acids. 2. The 3-Hydroxy-2-hydroxymethylpyrrolidines. (2001) *Journal of Org Chem.* 66(10): 3621 – 3626.