The Roles of Redox Active Cofactors in Catalysis:
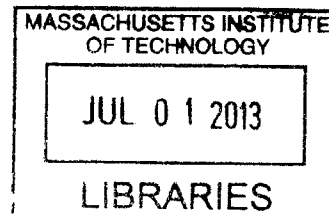## Structural Studies of Iron Sulfur Cluster and Flavin Dependent Enzymes

by

Peter John Goldman

A.B., Chemistry (2007)
Occidental College

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Signature of Author ....
Department of Chemistry
May 6, 2013

Certified by ....
Catherine L. Drennan
Professor of Chemistry and Biology
Howard Hughes Medical Institute Investigator and Professor
Thesis Supervisor

Accepted by..
Robert W. Field
Chairman, Departmental Committee on Graduate Students

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

Professor John M. Essigmann...
                                                             Committee Chairman
               William R. and Betsy P. Leitch Professor of Chemistry and Biological Engineering

Professor Catherine L. Drennan...
                                                            Research Supervisor
                                                Professor of Chemistry and Biology
                       Howard Hughes Medical Institute Investigator and Professor

Professor Thomas U. Schwartz...
                                                            Committee Member
                                                      Professor of Biology

The Roles of Redox Active Cofactors in Catalysis:
Structural Studies of Iron Sulfur Cluster and Flavin Dependent Enzymes

by

Peter John Goldman

ABSTRACT

Cofactors are highly prevalent in biological systems and have evolved to take on many functions in enzyme catalysis. Two cofactors, flavin adenine dinucleotide (FAD) and [4Fe-4S] clusters, were originally determined to aid in electron transfer and redox chemistry. However, additional activities for these cofactors continue to be discovered. The study of FAD in the context of rebeccamycin and staurosporine biosynthesis has yielded another role for this cofactor in the enzyme StaC. A homolog of this enzyme, RebC, uses its FAD cofactor in the oxidation of 7-carboxy-K252c. StaC also uses 7-carboxy-K252 as a substrate, but its reaction does not result in a redox transformation. Biochemical and X-ray crystallographic methods were employed to determine that, indeed, the role of FAD in the StaC system is not to catalyze redox chemistry. Instead, FAD sterically drives an initial decarboxylation event. Subtle differences in the active sites of RebC and StaC promote this redox neutral decarboxylation, by activating water for a final protonation step.

In another system, the characterization of the S-adenosyl-L-methionine (AdoMet) radical superfamily showed the versatility of these cofactors. In this superfamily, which includes over 40,000 unique sequences, [4Fe-4S] clusters are responsible for the initiation of radical chemistry. A recently described subclass of this superfamily, the dehydrogenases, require additional [4Fe-4S] cluster for activity. This requirement led to the hypothesis that these enzymes are catalyzing redox chemistry by directly ligating substrates to auxiliary (Aux) clusters. X-ray structures of 2-deoxy-$scyllo$-inosamine dehydrogenase (BtrN), required for the biosynthesis of 2-deoxystreptamine, and an anaerobic sulfatase maturating enzyme, anSMEcpe, which installs a required formylglycine posttranslational modification, refute this hypothesis. In these structures, substrate binding is distal from each enzymes' Aux clusters. However, the Aux cluster binding architecture shared between BtrN, anSMEcpe, and another AdoMet radical enzyme, MoaA, involved in molybdenum cofactor biosynthesis, suggests that the structural features will be a staple in the AdoMet radical superfamily, common to ~ 30% of the AdoMet radical reactions.

First off, thank you Cathy, for all of your time and effort teaching and mentoring me. It's been a pleasure learning from and working beside you for the past six years.

Thanks to everyone in the Drennan lab, past and present. To the folks who left before me, thank you for your patience and willingness to teach me crystallography. To everyone still around, thanks for keeping up a great lab culture and for helping make this thesis a reality.

Thank you, Allison, for your endless encouragement and enthusiasm. To my family and friends, thank you for your committed engagement in my life and your concern for my wellbeing. I'm very lucky to be surrounded and supported by such great people.

Thanks everyone. For everything.

Table of Contents

Chapter 3. X-ray Structure of an AdoMet Radical Activase Reveals an Anaerobic Solution for Formylglycine Posttranslational Modification   71

Chapter 1.

Introduction to the Flavin and Iron Sulfur Cluster Cofactors

I.I SUMMARY

Proteins are the workhorses of the cell and have evolved to carry out a variety of roles, including cellular signaling, gene regulation, and structural and catalytic functions. Catalytic proteins, called enzymes, are the cell's craftsmen, generating new molecules from a wide set of starting materials, allowing the cell to live. In the saying, "Life can be defined as one large set of coordinated chemical reactions," enzymes are the chemists. As in any laboratory, these chemists often need assistance to perform their duties. This dissertation will focus on two such cases. Chapter 2 will examine a step in the rebeccamycin and staurosporine biosynthetic pathways in which the activity of two enzymes is dependent on the molecule flavin adenine dinucleotide. The subjects of Chapters 3 and 4, two *S*-adenosyl-L-methionine radical dehydrogenase enzymes, require multiple iron sulfur clusters to function.

I.II INTRODUCTION

Most enzymes are able do chemistry using only their amino acid building blocks, which in certain conformations lower the activation energy of a chemical reaction. Some reactions, however, require additional components, i.e. cofactors, which are used when chemistry requires extra catalytic prowess, such as: a stronger nucleophile like thiamine pyrophosphate (vitamin B_1) and cyanocobalamin (vitamin B_{12}); functional group carriers like biotin (vitamin H) and coenzyme A (a derivative of vitamin B_5); or electron transfer agents. The latter are required for reduction/oxidation (redox) chemistry. Flavins, comprised of a common isoalloxazine tricyclic ring, and inorganic iron sulfur clusters are two cofactors whose traditional roles involve the transfer of electrons required for a variety of cellular processes. For example, they are prominently displayed in Complexes I, II, and III of the electron transport chain, where the funneling of electrons to different redox centers is coupled to the generation of a potential gradient. This gradient powers the mechanical turbine of the cell, ATP synthase, where it is converted to the chemical energy on which life relies (1). Flavins and iron sulfur clusters are thus central to a cell's ability to function. Disruption in their production or uptake causes multiple diseases in humans including Friedreich's ataxia and sideroblastic anemia for iron sulfur clusters (2) and trimethylaminuria and cardiovascular disease for flavins (3).

Cofactors can take many forms. The basic unit of flavin is vitamin B_2, or riboflavin, comprised of an isoalloxazine ring with a ribityl moiety attached at the N10 position. To assist

in their utilization, the flavin is tailored into two derivatives: flavin mononucleotide (FMN, riboflavin-5'-phosphate) and flavin adenine dinucleotide (FAD, riboflavin-5'-adenosine diphosphate, Figure I.1). Iron sulfur clusters exhibit even wider structural variations, from [2Fe-2S] clusters to [8Fe-8S] clusters, and can also be part of larger metallocomplexes, such as the [Ni-3Fe-4S] C-cluster of carbon monoxide dehydrogenase and the exquisite [Mo–7Fe–9S–C–homocitrate] cluster of nitrogenase (4-6). The most common cluster, however, is the [4Fe-4S] cluster. This cluster adopts a cubane structure, where each of the four iron atoms is tetracoordinate and ligated by three bridging sulfides from the cluster and a cysteine from a protein scaffold (Figure I.1). FAD and [4Fe-4S] clusters are traditionally used for one electron (FAD/[4Fe-4S] cluster) or two electron (FAD) transfers. In these roles, FAD can cycle between reduced, semiquinone, and oxidized states, while [4Fe-4S] clusters traditionally fluctuate between $[4Fe-4S]^{+1}$ and $[4Fe-4S]^{+2}$ states (Figure I.1). In this thesis, we use X-ray crystallography as our primary tool to study these two redox active cofactors in two nontraditional systems, summarized in the following sections.

I.III   Catalytic mechanism of the FAD dependent hydroxylase, RebC

While the role of FAD in electron transfer has been known since the early twentieth century, new roles for the cofactor are continually being found. The discovery that the cofactor could be used in halogenation has only been known for a little over a decade (7). This reaction is similar to another activity of flavoenzymes, hydroxylation. Both types of reactions are initiated by the reduction of molecular oxygen by FAD, resulting in the formation of a hydroperoxy species. This reactant turns the usually nucleophilic hydroxyl functional group into an electrophilic one, allowing the hydroxylation of electron rich substrates such as aromatics (Figure I.2). This hydroperoxy species can also react with a halide, similarly altering the nuceophilicity of halides, allowing for halogenation of electron rich substrates. This nucleophilic/electrophilic role reversal is central to the tailoring of traditionally incompatible substrates.

Katherine Ryan and Leah Blasiak, former graduate students in the Drennan group, along with members of the Walsh laboratory, were instrumental in the elucidation of the mechanisms of enzymes in the *reb* and *sta* pathways. These pathways are responsible for the biosynthesis of the natural products rebeccamycin and staurosporine. In the *reb* pathway, two FAD dependent enzymes, RebC and RebH, employ FAD for different purposes, one for the hydroxylation of chromopyrrolic acid (CPA) and the other for the chlorination of tryptophan,

respectively (8, 9). Both of these enzymes were structurally characterized in the Drennan group (10, 11). The *sta* pathway does not involve a chlorination reaction, but it contains an enzyme very similar to the CPA-modifying enzyme in the *reb* pathway, RebC. The mechanism of this enzyme, StaC, has been a subject of debate in the recent years because it does not catalyze an oxidation, and thus the role of FAD in this enzyme was enigmatic (12, 13). To understand the function of FAD in this enzyme, we decided to pursue mutation-driven biochemical investigations of both RebC and StaC. In addition, we solved structures of a 'StaC-like' enzyme, leading us to discover yet another role for FAD. These studies (14) are outlined in Chapter 2.

I.IV   The [4Fe-4S] cluster dependent AdoMet radical enzymes

In addition to electron transfer, [4Fe-4S] clusters can assist enzymatic catalysis by binding substrates. This function is seen in hydrolases (including aconitase and fumarase) and in the *S*-adenosyl-L-methionine (AdoMet, SAM) radical enzymes (4, 15). In its usual role, AdoMet is used as a methylating agent, resulting in methylated substrate and *S*-adenosylhomocysteine. When bound to a [4Fe-4S] cluster, however, an inner sphere electron transfer event triggers the homolysis of a carbon – sulfur bond in AdoMet, generating methionine and a 5'-deoxyadenosyl radical (5'-dA•) (Figure I.3). This radical species is a potent oxidant that is able to abstract hydrogens from unactivated positions, generating substrate radicals. After hydrogen abstraction, the substrate radical can go down multiple paths, inducing a wide variety of transformations (16). In these systems, the AdoMet is either reformed in the reaction and used catalytically, or consumed stoichiometrically (Figure I.3).

The basis for identifying AdoMet radical enzymes from sequence databases has been a three cysteine motif, CX_3CX_2C, responsible for coordinating the three non-AdoMet iron ligation sites of the [4Fe-4S] cluster (17). This sequence searching led to the identification of nearly 50,000 unique AdoMet radical enzymes, as assigned by the Structure Function Linkage Database (http://sfld.rbvi.ucsf.edu). Between reaction classes, these enzymes have little to no sequence similarity outside the three cysteine motif (17). Combined with the high modularity of these enzymes due to their widely varied reactions and substrates, this lack of similarity has made sequence based functional predictions of the AdoMet radical enzymes difficult. Surprisingly, however, these enzymes have very similar three dimensional structures (18, 19).

The first structural characterizations of AdoMet radical enzymes came just two years after the superfamily was initially described (17). The structures of BioB and HemN, two enzymes with just 9% sequence identity, overlaid surprisingly well, with a root-mean-square

deviation (RMSD) of 3.6 Å (20, 21). This structural alignment shows that both enzymes contain a partial beta barrel at their core, reminiscent of a triose-phosphate isomerase (TIM) $(\beta/\alpha)_8$ barrel. These partial TIM barrels, or $(\beta/\alpha)_6$ structures, overlay even better, with an RMSD of 2.6 Å for BioB and HemN (Figure I.4). Their [4Fe-4S] cofactors and bound AdoMet molecules are also nearly superimposable, with the three cysteines in the enzymes' CX_3CX_2C motifs residing in a loop between $\beta1$ and $\alpha1$ in both enzymes. Here, AdoMet is bound to the available, 'unique', iron site of the [4Fe-4S] cluster via the amino and carboxyl groups of the AdoMet methionine moiety, matching available spectroscopic evidence for the coordination unique iron position (15, 22, 23).

Since the structural characterizations of BioB and HemN reported a decade ago, fourteen additional X-ray structures of AdoMet radical enzymes have been released by the Protein DataBank. Of these, twelve have AdoMet radical folds nearly identical to those found in BioB and HemN. This conservation led to the description of many AdoMet binding and structural motifs that have roots in the original structures of BioB and HemN (18, 19, 24). Two additions to the structural characterized AdoMet radical enzymes, anSMEcpe and BtrN, are presented in Chapters 3 and 4 of this dissertation. This work provides a further description of these motifs. In addition, these two structures, along with the previously characterized MoaA enzyme, contain auxiliary [4Fe-4S] cluster binding domain substructures which could be present in up to 30% of AdoMet radical enzymes. As discussed throughout this thesis, these auxiliary clusters seem to serve the traditional [4Fe-4S] cluster function, providing a conduit for the transfer of electrons. Chapter 5 includes a discussion of future directions and predictions for this interesting subfamily and deviations to the AdoMet radical fold motif presented by BtrN and another enzyme, QueE.

I.V   CONCLUSION

Interestingly, the catalytic functions of FAD and [4Fe-4S] clusters combine the traditional use of the cofactors (electron transfer) with new roles (generation of hydroperoxy species and AdoMet homolysis). These ubiquitous cofactors continue to surprise biochemists. The struggle to find and characterize exceptions to fundamental theories in each field continues to drive these studies, and in all cases, touches upon our ever expanding understanding of Nature's ingenuity and complexity.

Figure I.1. Physiologic redox states of the cofactors (A) FAD and (B) [4Fe-4S] cluster.

A.

Adenine monophosphate

Ribityl-5'-phosphate

Isoalloxazine ring

reduced (hydroquinone)

semiquinone

oxidized (quinone)

B.
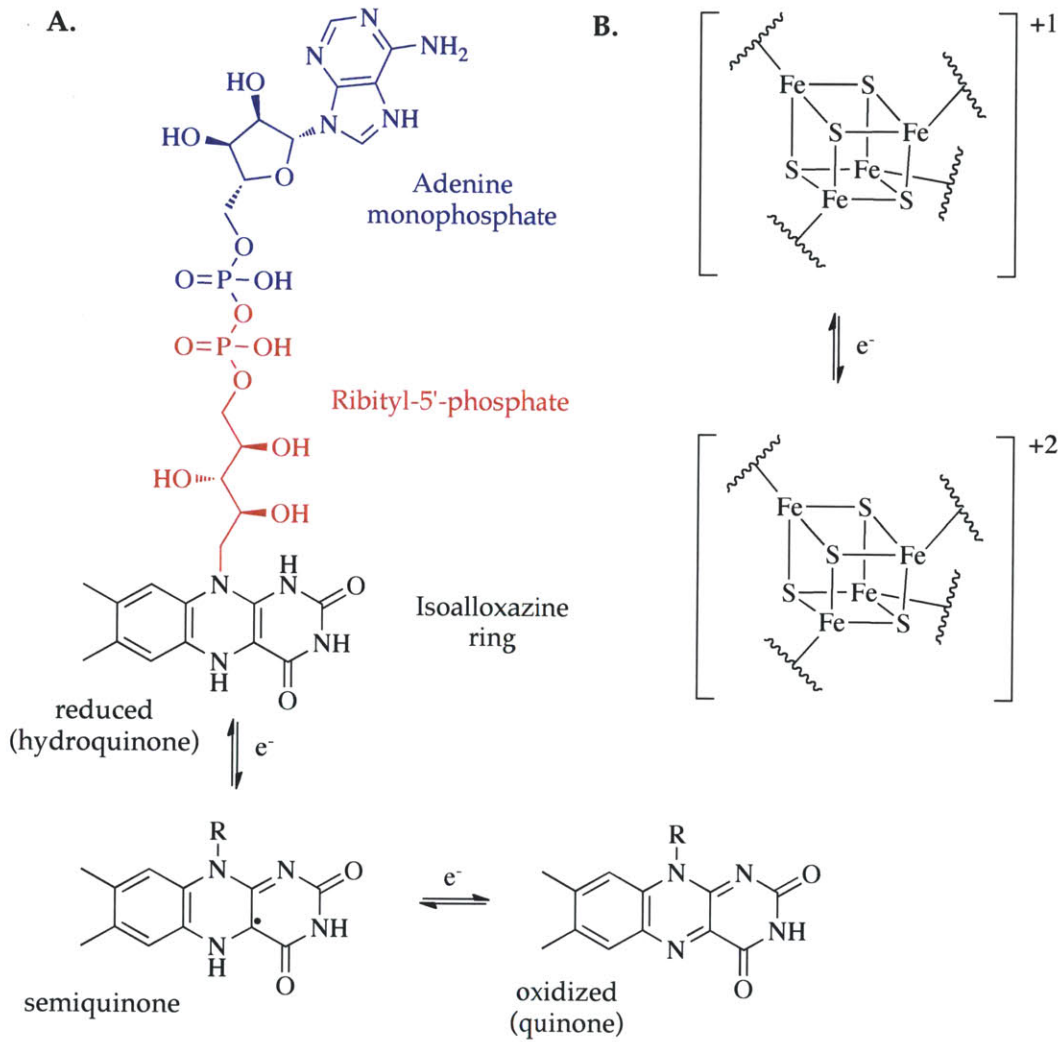
Figure I.2. Activation of molecular oxygen for the hydroxylation of nucleophilic positions.
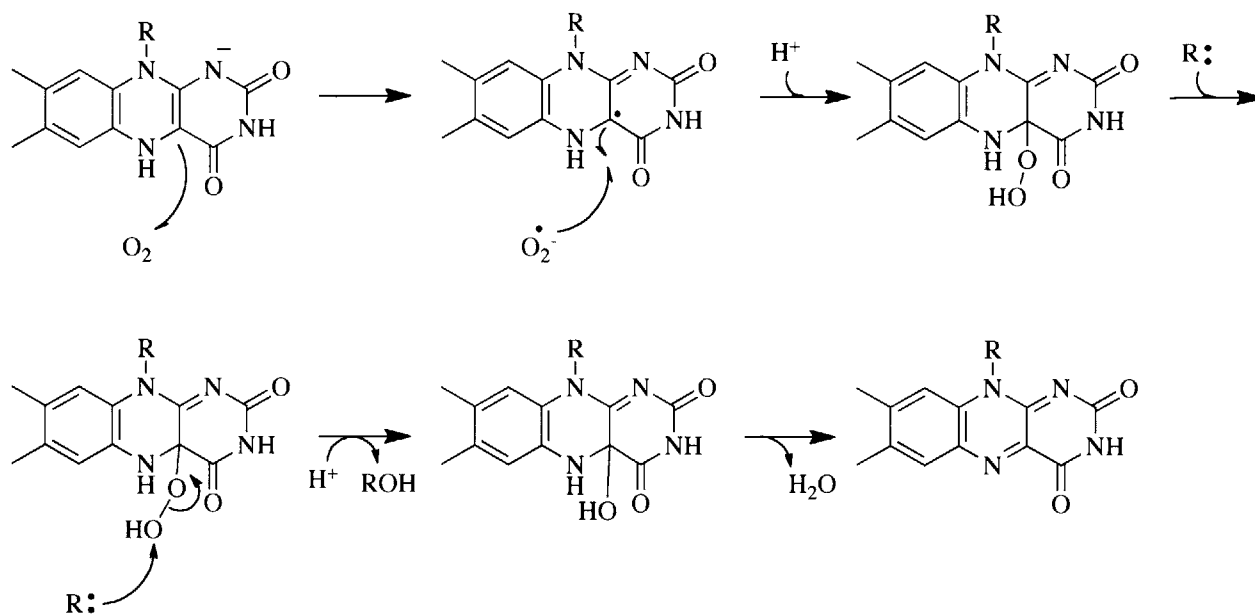
Figure I.3. Generation and use of 5'-dA•. (A) AdoMet binds to the unique iron site of a [4Fe-4S]$^{2+}$ cluster. Upon reduction of the cluster to 1+ state and electron transfer, AdoMet is homolytically cleaved, forming methionine and a 5'-deoxyadenosyl radical (5'-dA•). AdoMet can either be used (B) catalytically or (C) stoichiometrically.

A.

B.　　5'-dA• + SubstrateH → 5'-dAH + Substrate• → Product•

Product• + 5'-dAH + Met → ProductH + AdoMet

C.　　5'-dA• + SubstrateH → 5'-dAH + Substrate• → 5'-dAH + Product
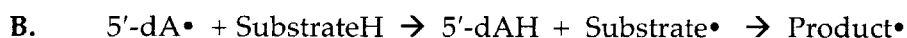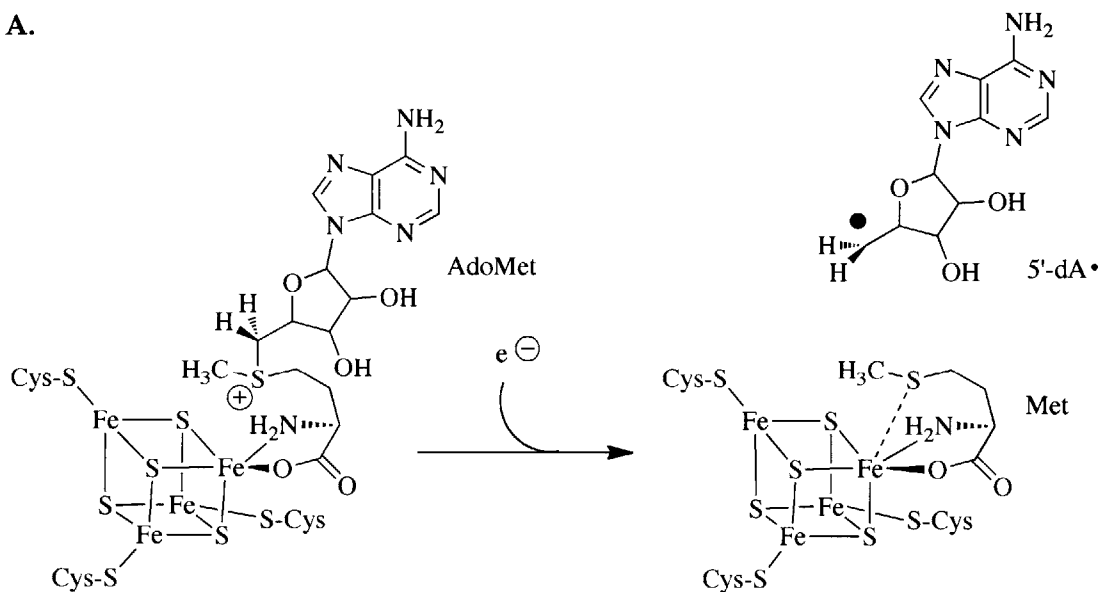
Figure I.4. The AdoMet radical folds of BioB and HemN. The architecture of (*A*) BioB (PDB ID 1R30) and (*B*) HemN (PDB ID 1OLT) with helices in red, beta sheets in yellow and loops in green; AdoMet and [4Fe-4S] cluster in sticks; AdoMet in grey carbons. (*C*) An alignment of the BioB and HemN (β/α)$_6$ partial TIM barrel folds (residues 41-223 and 53-241, respectively); RMSD 2.6 Å.

A.

B.

C.
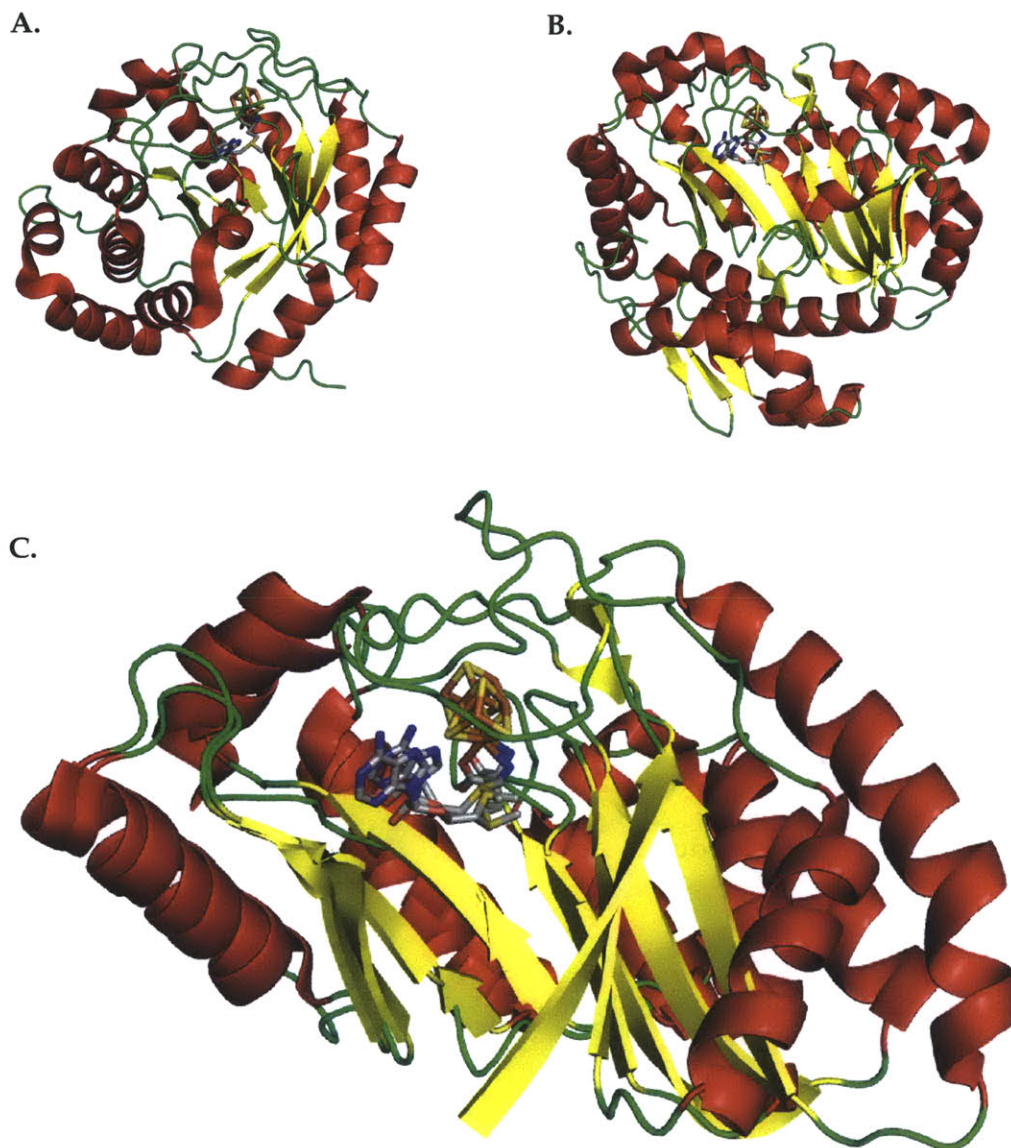
I.VI REFERENCES

1. Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem* 54:1015-1069.

2. Rouault TA & Tong WH (2008) Iron-sulfur cluster biogenesis and human disease. *Trends Genet* 24(8):398-407.

3. Powers HJ (2003) Riboflavin (vitamin B-2) and health. *Am J Clin Nutr* 77(6):1352-1360.

4. Johnson DC, Dean DR, Smith AD, & Johnson MK (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annual Review of Biochemistry* 74:247-281.

5. Lancaster KM, *et al.* (2011) X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* 334(6058):974-977.

6. Drennan CL, Heo J, Sintchak MD, Schreiter E, & Ludden PW (2001) Life on carbon monoxide: X-ray structure of Rhodospirillum rubrum Ni-Fe-S carbon monoxide dehydrogenase. *Proc Natl Acad Sci USA* 98(21):11973-11978.

7. Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, & Walsh CT (2006) Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem Rev* 106(8):3364-3378.

8. Yeh E, Garneau S, & Walsh CT (2005) Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc Natl Acad Sci USA* 102(11):3960-3965.

9. Howard-Jones AR & Walsh CT (2006) Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. *J Am Chem Soc* 128(37):12289-12298.

10. Yeh E, Blasiak LC, Koglin A, Drennan CL, & Walsh CT (2007) Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* 46(5):1284-1292.

11. Ryan KS, *et al.* (2007) Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC. *Proc Natl Acad Sci USA* 104(39):15311-15316.

12. Asamizu S, Shiro Y, Igarashi Y, Nagano S, & Onaka H (2011) Characterization and Functional Modification of StaC and RebC, Which Are Involved in the Pyrrole Oxidation of Indolocarbazole Biosynthesis. *Biosci Biotech Bioch* 75(11):2184-2193.

13. Groom K, Bhattacharya A, & Zechel DL (2011) Rebeccamycin and staurosporine biosynthesis: Insight into the mechanisms of the flavin-dependent monooxygenases RebC and StaC. *Chembiochem* 12(3):396-400.

14. Goldman PJ, *et al.* (2012) An unusual role for a mobile flavin in StaC-like indolocarbazole biosynthetic enzymes. *Chem Biol* 19(7):855-865.

15. Krebs C, Broderick WE, Henshaw TF, Broderick JB, & Huynh BH (2002) 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):912-913.

16. Frey PA, Hegeman AD, & Ruzicka FJ (2008) The Radical SAM Superfamily. *Crit Rev Biochem Mol Biol* 43(1):63-88.

17. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, & Miller NE (2001) 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):1097-1106.

18. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.

19. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.

20. Layer G, Moser J, Heinz DW, Jahn D, & Schubert WD (2003) Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J* 22(23):6214-6224.

21. Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, & Drennan CL (2004) Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* 303(5654):76-79.

22. Walsby CJ, Ortillo D, Broderick WE, Broderick JB, & Hoffman BM (2002) 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):11270-11271.

23. Walsby CJ, *et al.* (2002) 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):3143-3151.

24. Nicolet Y & Drennan CL (2004) AdoMet radical proteins--from structure to evolution--alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* 32(13):4015-4025.

Chapter 2.

An Unusual Role for a Mobile Flavin in StaC-like Indolocarbazole Biosynthetic Enzymes

Authors:

Peter J. Goldman,[1,7] Katherine S. Ryan,[2,4,7] Michael J. Hamill,[5] Annaleise R. Howard-Jones,[6] Christopher T. Walsh,[6] Sean J. Elliott,[5] and Catherine L. Drennan[1,2,3*]

[1]Department of Chemistry, [2]Department of Biology, and the [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA
[4]Current address: Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1 Canada
[5]Department of Chemistry, Boston University, Boston, Massachusetts 02215 USA
[6]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115 USA
[7]These authors contributed equally to this work

II.I SUMMARY

The indolocarbazole biosynthetic enzymes StaC, InkE, RebC, and AtmC mediate the degree of oxidation of chromopyrrolic acid on route to the natural products staurosporine, K252a, rebeccamycin, and AT2433-A1, respectively. Here we show that StaC and InkE, which mediate a net 4-electron oxidation, bind FAD with a micromolar K_d, whereas RebC and AtmC, which mediate a net 8-electron oxidation, bind FAD with a nanomolar K_d, while displaying the same FAD redox properties. We further create RebC-10x, a RebC protein with ten StaC-like amino acid substitutions outside of previously characterized FAD binding motifs and the complementary StaC-10x. We find that these mutations mediate both FAD affinity and product specificity, with RebC-10x displaying higher StaC activity than StaC itself. X-ray structures of this StaC catalyst identify the substrate of StaC as 7-carboxy-K252c and suggest a unique mechanism for this FAD-dependent enzyme.
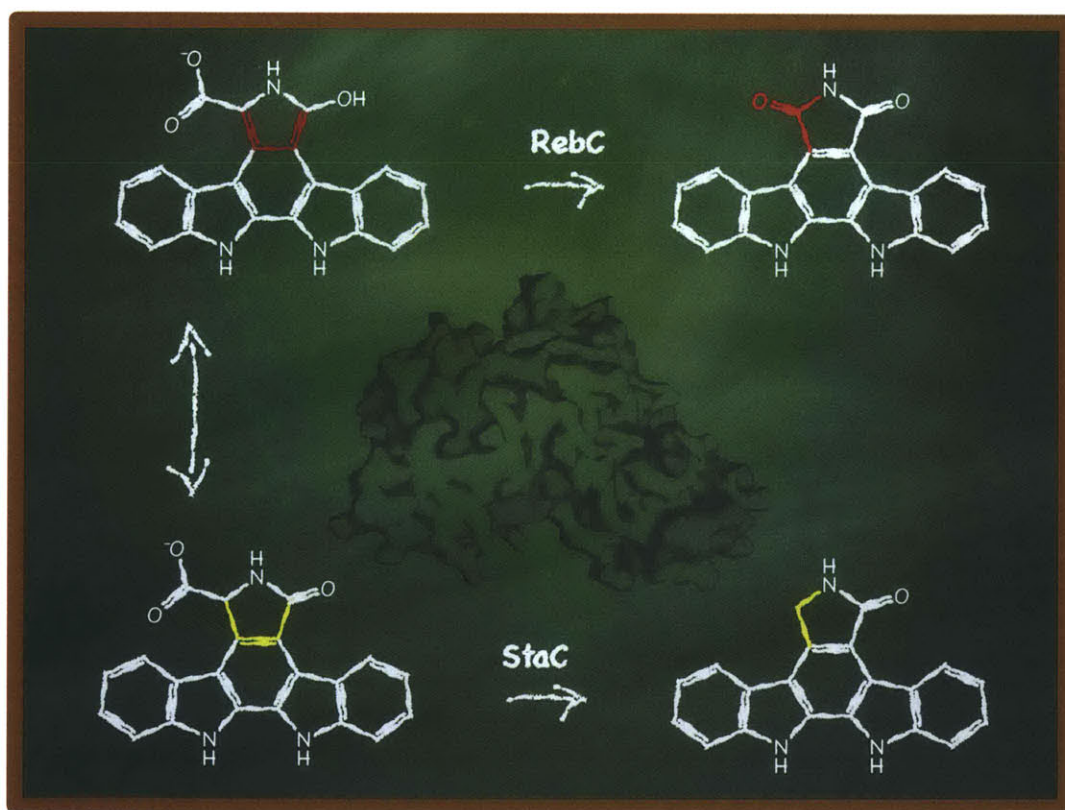
II.II   INTRODUCTION

Indolocarbazoles, a subset of the L-tryptophan derived bisindole class of alkaloid natural products, include a variety of molecules of pharmaceutical interest (1). Isolated from *Streptomyces* and other soil- and marine-dwelling actinomycete bacteria (2, 3), staurosporine (Figure II.1.A) has no assigned native function but has proven to be a potent protein kinase inhibitor (4) with an analog (7-hydroxy-staurosporine, also known as UCN-01) previously in clinical trials as an anti-cancer agent (5-7). Rebeccamycin (Figure II.1.A), isolated from *Lechevalieria aerocolonigenes*, shares a very similar aglycone scaffold with staurosporine and an analog of this compound, becatecarin, is also in clinical trails afforded by its antitumor activity (8, 9). However, unlike staurosporine, rebeccamycin analogs inhibit DNA replication by stabilizing DNA-topoisomerase I complexes (10), showing the therapeutic diversity of indolocarbazole compounds. To exploit the full medicinal potential of indolocarbazoles and to generate a more diverse array of pharmaceutically active compounds, a better understanding of their biosynthetic pathways and the key branch points is desirable.

The biosynthetic pathways of staurosporine, rebeccamycin, and other indolocarbazoles involve the oxidation and subsequent coupling of two tryptophan molecules to generate chromopyrrolic acid (CPA) (Figure II.1.B). At this step, the pathways diverge to generate the diverse aglycone scaffolds characteristic of indolocarbazoles. In the staurosporine pathway, CPA is converted to K252c in a net 4-electron oxidation, and then further tailored, whereas CPA is converted to arcyriaflavin A via a net 8-electron oxidation in the rebeccamycin biosynthetic pathway (Figure II.1). Both oxidations involve two enzymes, the first of which, StaP or RebP, are cytochrome P450 enzymes that catalyze aryl-aryl coupling reactions to generate a reactive intermediate and can be used interchangeably between the two pathways (11). The second enzyme of each pathway, StaC or RebC, is thought to intercept this intermediate and enable its conversion to the corresponding aglycone (12). For StaC, the reaction yields an asymmetric molecule with a carbonyl group at the C-5 position and a fully reduced C-7 carbon (numbering in Figure II.1.A), while the product of RebC is a symmetric molecule, with carbonyl groups at both the C-5 and C-7 positions. Without either StaC or RebC in the reaction mix, a variety of products are generated with varying oxidation states at the C-7 position; however, if either enzyme is included, a single product is observed (12).

StaC and RebC share 65% sequence identity and both enzymes contain three motifs typical of flavin hydroxylase proteins (Figure II.2) (12-14). Surprisingly though, only RebC co-purifies with FAD in *E. coli* recombinant expression systems. Nonetheless, StaC is capable of converting FAD to $FADH_2$ using NAD(P)H (12). Two protein homologues of RebC and StaC,

InkE and AtmC (Figure II.2), are involved in separate indolocarbazole biosynthetic pathways. Both contain the same three motifs and the enzymes share 56% sequence identity with one another. InkE is involved in the biosynthesis of K252a (Figure II.1.A) (15), a molecule that has a carbonyl group at the C-5 position but a fully reduced C-7 carbon. The aglycone of K252a is likely generated through a largely similar pathway as the staurosporine aglycone, with InkE (like StaC) finalizing a net 4-electron oxidation of CPA (Figure II.1.B). By contrast, AtmC is involved in the biosynthesis of AT2433-A1 (Figure II.1.A) (16), a molecule that has carbonyl groups at the both the C-5 and C-7 positions. The aglycone of AT2433-A1 is likely generated through a highly similar pathway as the rebeccamycin aglycone, with AtmC (like RebC) mediating a net 8-electron oxidation of CPA (Figure II.1.B).

Here we investigate the role of FAD and the enzyme mechanism of RebC- and StaC-like biosynthetic enzymes. Using isothermal titration calorimetry (ITC) to determine FAD dissociation constants for StaC, RebC, InkE, and AtmC, we find a correlation between FAD affinity and the reaction catalyzed, with tighter FAD affinity linked with RebC/AtmC-like activity and weaker with StaC/InkE-like activity. To investigate whether mutations that alter FAD affinity also alter the type of reaction catalyzed, we use the structure of RebC (13) in combination with comparative sequence analysis of the enzyme family (RebC, StaC, AtmC, and InkE) to generate a RebC protein with ten amino acid substitutions, called RebC-10x, and the complementary StaC-10x protein (Table II.1). Excitingly, we find that RebC-10x shows a decrease in FAD affinity (although not to StaC levels) and exhibits strong StaC-like activity, while StaC-10x shows an increase in FAD affinity (although not up to RebC levels) and is a weak RebC-like catalyst. To probe the molecular basis for these enzyme activity conversions, FAD redox potentials are measured and the structure of RebC-10x is determined in native, substrate-bound and product-bound states. These data, along with recent site-directed mutagenesis studies of singly and doubly mutated RebC and StaC proteins (17, 18), suggest the structural basis for the differential catalytic activities of these proteins.

II.III   RESULTS

The FAD binding affinity correlates with the reaction catalyzed for wild-type enzymes

StaC, RebC, AtmC, and InkE were purified recombinantly without addition of exogenous riboflavin to the media. As reported earlier, StaC purifies without bound FAD, whereas RebC purifies with bound FAD (12). Using a new cell line for protein expression (Rosetta™ (DE3) 2 pLysS cells from Novagen), we were able to increase the amount of FAD co-

purified with RebC to ~68%, from ~33% reported previously, when RebC was purified from BL21(DE3) cells (12). Furthermore, we find that AtmC, which catalyzes a RebC-like reaction, purifies with bound FAD, while InkE, which catalyzes a reaction similar to StaC, purifies without bound FAD.

To determine the FAD dissociation constants for these four proteins, we generated the deflavinated (apo) forms and performed isothermal titration calorimetry experiments (ITC). The removal of the FAD cofactor was unnecessary in the case of StaC and InkE, as each of these proteins purifies without bound FAD. For RebC and AtmC, deflavination was accomplished by washing the protein bound to the metal affinity column with a buffer containing 2 M KBr and 2 M urea (19). We find that the dissociation constants (K_d's) of StaC, RebC, InkE, and AtmC for FAD are 15 ± 2 µM, 20 ± 12 nM, 59 ± 48 µM, and 73 ± 28 nM, respectively (Table II.2 and Figures 1.3.A-F). While RebC and AtmC, both of which facilitate a net 8-electron oxidation of CPA (12, 16), have FAD K_d's in the low nanomolar range, StaC and InkE, both of which facilitate a net 4-electron oxidation of CPA (12, 15), have FAD K_d's in the micromolar range (Table II.2). These data suggest that, among this group of enzymes, there is a correlation between, on one hand, tight (nanomolar) binding of FAD and mediation of a net 8-electron oxidation of CPA and, on the other hand, substantially weaker (micromolar) binding of FAD and mediation of a net 4-electron oxidation of CPA.

StaC-10x and RebC-10x have similar, intermediate affinities for FAD

Because of the close sequence similarity of RebC, StaC, AtmC, and InkE, and because the structure of RebC is available (13), the identity of residues likely to mediate the differences in enzymatic activity and FAD affinity between the two groups of enzymes could be readily predicted (Figure II.2). With few exceptions, sequence alignment reveals key residues that have one identity in both RebC and AtmC and another identity in both StaC and InkE (Figure II.2). Using the RebC crystal structure to filter out residues distant from the expected FAD-binding and active sites, ten RebC residues and ten StaC residues were chosen for complementary mutation to the amino acid found in the homologous enzyme, resulting in our RebC-10x and StaC-10x constructs (Table II.1).

To assess their FAD binding affinity, these new constructs were expressed and purified for ITC experiments. Like RebC and AtmC, both RebC-10x and StaC-10x co-purify with FAD. RebC-10x was deflavinated in the same manner as RebC and AtmC, and its K_d for FAD was determined to be 600 ± 55 nM (Figure II.3.E, Table II.2). Unlike RebC-10x, however, it was not

27

possible to produce apo-StaC-10x (See Methods). Therefore as isolated StaC-10x, which co-purifies with 33% FAD occupancy, was used in ITC experiments to yield an approximate K_d of FAD for StaC-10x of 560 ± 120 nM (Figure II.3.F, Table II.2). To understand the relationship of this value to the true K_d, we determined the K_d of FAD for as-purified RebC without removal of FAD (RebC co-purifies with ~68% FAD in our expression system). The K_d of RebC for FAD determined in this manner is 31 nM, a value within the range of values determined using fully deflavinated protein (20 ± 12 nM). Therefore, although the K_d calculated for StaC-10x for FAD is imprecise, we assume that the value is on the order of magnitude of the true K_d value. Thus, by targeting ten amino acids that may be involved in binding FAD, we have altered both StaC's and RebC's affinity for FAD, suggesting successful identification of residues important for modulating FAD affinity.

RebC-10x and StaC-10x have almost identical redox potentials to RebC

To determine whether the amino acid replacements had an effect on the FAD midpoint potential, reduction potentials of RebC-10x and StaC-10x were determined and compared to the previously reported reduction potential of RebC, –179 mV (13). Using the xanthine/xanthine oxidase reduction assay (20, 21), the reduction potential of FAD-bound RebC-10x was found to be –178 mV ± 5 (Figure II.4.A), identical within experimental error to that of FAD-bound RebC. Interestingly, the reduction potential of the StaC-10x FAD is also the same – 178 ± 2 mV (Figure II.4.B). These results show that the mutant proteins have adopted or retained a flavin cofactor redox environment similar to that found in RebC.

The 10x mutant proteins have altered activities

Activity assays using a previously described method (12) were performed to investigate the chemical reactivities of both 10x variants. Because the substrate(s) of StaC and RebC have never been isolated, this assay uses the upstream enzyme StaP and its known substrate CPA together with an electron source provided by spinach ferredoxin, flavodoxin NADP$^+$ reductase, and NADPH to generate the reactive intermediate that is the putative substrate(s) for RebC and StaC. The reaction products are then separated and analyzed via HPLC. In the presence of StaP alone, the reaction yields multiple products, oxidized to different degrees, including K252c (a 4-electron oxidation product), 7-hydroxy-K252c (a 6-electron oxidation and the major product), and arcyriaflavin A (an 8-electron oxidation product). The addition of either RebC or StaC to the

assay directs the reaction to a single major product, arcyriaflavin A for RebC and K252c for StaC (12).

Interestingly, when RebC-10x is added to the assay, no production of the RebC product arcyriaflavin A is observed. Instead, RebC-10x addition results in production of the StaC product K252c, with a rate that is ~3-fold faster than the rate of the StaC-catalyzed reaction (based on authentic samples and standard curves) (Table II.3 and Figure II.5). StaC-10x behaves like RebC to the extent that it produces arcyriaflavin A as its major product and has lost nearly all of its activity for the normal StaC product K252c. However, it is not a strong RebC catalyst, generating arcyriaflavin A at rates that are ~8-fold lower than that of wild-type RebC (Table II.3 and Figure II.5).

Structures of RebC-10x display the key characteristics of a flavin hydroxylase and explain the molecular basis for decreased FAD affinity

Crystal structures spanning three different states of the RebC-10x reaction (native, substrate-, and product-bound) have been determined in an effort to provide a molecular explanation for the decrease in FAD affinity and gain in StaC-like function. A single crystal structure of K252c-soaked RebC-10x with two molecules in the asymmetric unit has afforded a view of both native and product-bound RebC-10x states at 2.33 Å resolution, while soaking a RebC-10x crystal with StaP substrate CPA (see below) has provided a substrate-bound structure at 2.76 Å resolution (see Table II.4 for data collection and refinement statistics). Structures were solved by rigid body refinement against the wild-type RebC structure, with mutations confirmed by $F_o–F_c$ difference density. As expected, RebC-10x adopts the general fold of flavin-dependent hydroxylases (22) with FAD-, substrate-binding and thioredoxin-like domains (Figure II.6.A). The root mean squared deviation (RMSD) of mainchain atoms between RebC-10x and RebC structures are less than 1 Å, indicating no major conformational changes occur as a result of the ten mutations (Figure II.6.A). In addition, the mutations do not appear to cause local instability, as B-factors of the mutated sites and their surroundings are very similar when compared to the structure of wild type RebC (Figure II.7). Other RebC features retained in RebC-10x, such as the 'mobile flavin' and the 'melting helix,' are revealed by the structures generated through the soaking of RebC-10x with StaC product, K252c. The 'out' position of the mobile flavin is observed in the molecule in the asymmetric unit that contains no bound K252c. This FAD orientation is associated with NADPH reduction due to the afforded solvent accessibility (Figure II.6.B) (13, 23, 24). The 'in' position or 'catalytic orientation' of the mobile

flavin is found in the molecule that has density for K252c. Comparison of these two structures also shows that like RebC, residues 354-363 (termed the 'melting helix') are disordered when the active site is empty and become ordered when the active site is full (Figure II.6). This 'melting helix' region is believed to be the entry point for substrate (13).

Differences between RebC and RebC-10x are also observed. As expected, multiple protein-FAD interactions are lost when the RebC-10x structure is compared to RebC. These include the loss of van der Waals interactions between Q37 and T38 and the FAD adenosine moiety when these residues are replaced by A37 and G38, resulting in a more solvent accessible cofactor (Figure II.8.A); a hydrogen bond loss from E36 to the FAD ribose when E36 is replaced with shorter D36 (Figure II.8.A); a salt-bridge loss from R46 to the FAD α-phosphate when R46 is replaced with shorter K46 (Figure II.8.D); and another hydrogen bond loss from R239 of RebC to FAD O4 in the 'out' position, when R239 is replaced with N239 in RebC-10x (Figure II.8.E).

CPA-soaked RebC-10x structure reveals putative StaC substrate

A structure of RebC-10x bound to its putative substrate was obtained by incubating RebC-10x crystals with StaP substrate, CPA, for one week. Previously, the putative substrate for RebC was 'trapped' by allowing CPA to spontaneously degrade in the presence of RebC crystals. The crystallized RebC protein selectively bound and stabilized its substrate (7-carboxy-K252c) in its active site, allowing visualization (13). Using this same approach with the StaC catalyst RebC-10x, indolocarbazole density was discovered in the active site of RebC-10x (Figure II.9). Since the identity of the StaC (and thus RebC-10x) substrate is not known, a number of indolocarbazoles were modeled into the electron density, including aryl-aryl coupled CPA, arcyriaflavin A, K252c, and the enol tautomer of 7-carboxy-K252c, the molecule identified in the previous RebC CPA soak (Figure 1.B). Refinement of these molecules results in significant positive and negative difference density peaks (Figure II.9.A-D), indicating that they do not accurately describe the electron density. However, when the *S*-keto tautomer of 7-carboxy-K252c is modeled, little or no residual F_o-F_c difference density appears (Figure II.9.E) and the refined structure is the best fit to omit density maps (Figure II.9.G-H). Thus, our data support the presence of the *S*-keto tautomer of 7-carboxy-K252c in the RebC-10x active site, and suggest that this molecule is likely the substrate for StaC.

Interestingly, while RebC binds 7-carboxy-K252c with its carboxyl group pointed toward R230 and its carbonyl in a more hydrophobic pocket (Figure II.10.A), RebC-10x binds the *S*-keto tautomer of 7-carboxy-K252c in two orientations that differ by 180° (Figure II.10.B-C).

One molecule in the asymmetric unit, molecule A, binds substrate in the same orientation as RebC (Figure II.10.B), while in molecule B, the substrate sits in an alternate orientation, with the carbonyl pointing toward R230 and the carboxyl group near the hydrophobic pocket (Figure II.10.C). Notably, FAD is missing in these structures. During the week-long CPA soak used to generate this crystal structure, FAD, which binds more weakly to RebC-10x than RebC, dissociates, and a loop containing residues 45 – 48 shifts to occupy its site, repositioning P45 such that it now interacts with W276 (Figure II.10.A-C). The absence of FAD yields a larger active site, which can accommodate both orientations of the 7-carboxy-K252c. Modeling FAD in the 'in' conformation into the active site predicts a tight fit with substrate in either orientation, suggesting that FAD may be in an 'out' conformation prior to decarboxylation or that FAD moving 'in' might prompt decarboxylation. Importantly, regardless of FAD position, both 7-carboxy-K252c molecules are bound to RebC-10x in the S-keto tautomer form, while the enol tautomer of 7-carboxy-K252c was found in RebC. With active sites of StaC-catalyst RebC-10x so similar in sequence to RebC, it is interesting to consider the molecular basis for this preferential binding.

RebC and RebC-10x active site comparisons

Although mutations were designed to alter the affinity of FAD for the enzyme and most reside in the FAD binding pocket, some residues (like G48S, T241V, F216V, R239N) are also near the substrate binding site and are likely to influence both substrate binding and reactivity (Figure II.8.C-F). While the loss of FAD causes a repositioning of the FAD binding loop (residues 39–46), most of the protein backbone in the substrate-binding site of RebC-10x is identical to that of RebC, allowing direct comparison of substrate-bound RebC10x to substrate-bound RebC. A key difference between the two is the removal of salt bridge and hydrogen bonding interactions between R239 and the C-7 carboxyl group and NH-6 position of the RebC planar substrate (Figure II.10.D). In RebC-10x, N239 is incapable of interacting with either the bound RebC-10x keto substrate, or the overlaid enol tautomer from the RebC structure (Figure II.10.D).

Another difference between the two active sites is the positioning of another arginine, R230. The residue, present in both RebC and RebC-10x, sits lower in the RebC-10x active site, apparently due to the combination of the G48S and F216V mutations (Figure II.10.D). In particular, the substitution of S for G48 sterically blocks the RebC R230 position and V for F216 creates room for this residue to move down. This 'lower' position of R230 in RebC-10x disrupts

the favorable interaction observed in RebC between the guanidinium group of R230 and the 7-carboxyl of substrate in the enol tautomeric form. In RebC, the distance (2.6 Å) as well as the geometry of the interaction should serve to stabilize this orientation of the 7-carboxyl group and thus stabilize the enol form of the substrate (Figure II.10.D). In contrast, the position of R230 observed in RebC-10x has suboptimal geometry to maintain this interaction, and we observe instead, a new position of the 7-carboxyl moiety and a keto tautomeric form of substrate. This altered orientation of R230 in RebC-10x no longer interacts with substrate, but rather appears to be involved in creation of a new water binding site not observed in RebC. When water is modeled into observed positive F_o-F_c electron difference density adjacent to R230, the difference electron density disappears and the B-factor for this new water in molecules A and B, 65.0 and 58.6 Å^2, are consistent with the average B-factors for protein atoms (62.7 Å^2) and 7-carboxy-K252c atoms (66.1 Å^2), indicating that water is a good fit for this density. Following refinement, the water is located 3.1 and 3.4 Å from C-7 position of the RebC-10x substrate in chains A and B, respectively (Figure II.9.A and II.10.E), a position that suggests a catalytic role (see discussion). This water binding site is conserved in both the native and K252c bound RebC-10x molecules. Overall, the active sites of RebC and RebC-10x are similar; the bound products and substrates can be superimposed. However, small changes in side chain positions results in different tautomers of substrate and changes to bound water positions.

II.IV   DISCUSSION

While the substrates for RebC and StaC can convert to the respective products spontaneously, the presence of these enzymes directs the reactivity toward a single product. In this sense, RebC and StaC are babysitting enzymes, ensuring that the appropriate reaction goes forward, but just as importantly, preventing unwanted reactions. With such high sequence similarly between RebC and StaC, understanding how these enzymes catalyze their differential reactions has been a point of considerable interest in the natural product field.

For RebC, biochemical and structural analysis helped elucidate the nature of the RebC substrate as 7-carboxy-K252c (13), indicating that conversion to product requires both a decarboxylation and oxidation at C-7. With structural analysis firmly placing RebC in the FAD monooxygenase class of enzymes, it has been proposed that RebC catalyzes both the decarboxylation and oxidation of substrate using FAD-dependent hydroxylase activity (13, 25). In contrast StaC, which purifies in a deflavinated state, does not require oxidation activity, raising the question of whether StaC requires FAD for activity. In this work, we find a clear

correlation between FAD-binding affinities and the reaction catalyzed, with tighter (nanomolar) binding associated with enzymes that mediate a net 8-electron oxidation of CPA and substantially weaker (micromolar) binding associated with the mediation of a net 4-electron oxidation of CPA. This trend holds for both the wild-type and mutant proteins. However, we find that RebC-10x, which binds its FAD more tightly than StaC by three orders of magnitude, is actually a better StaC catalyst than StaC itself, indicating that truly weak binding of FAD is not required for StaC-like chemistry. In contrast, StaC-10x, which binds FAD more weakly than RebC by 30-fold, is a poor RebC catalyst. Collectively, these data imply that both StaC and RebC are FAD-dependent enzymes in which the tight binding of FAD is more important to the facilitation of RebC-like chemistry. Further support for the identification of StaC as an FAD-dependent enzyme comes from Onaka and co-workers, who showed that StaC activity is substantially diminished when regions of the GXGXXG FAD binding motif are mutated (17).

With FAD independence ruled out as the source of StaC's disparate reactivity, we further considered if variations in substrate use and/or FAD-based mechanisms could be the cause. While it is difficult to envision a FAD monooxygenase mechanism that could convert 7-carboxy-K252c to K252c, Onaka and co-workers argued that such a mechanism could be invoked for converting a substrate such as 5-deoxo-7-carboxy-K252c to form K252c (17). To investigate the nature of the StaC substrate, we employed the same crystallographic trick that was successful for RebC (13): screening CPA degradation products using enzyme crystals to determine which molecule selectively binds. Interestingly, we again find 7-carboxy-K252c in the active site, implying RebC and StaC do indeed share a common substrate. The fact that 7-carboxy-K252c is found in the structures of both RebC and the StaC-like catalyst RebC-10x is consistent with the interchangeable use of the preceding enzymes (StaP and RebP).

Given that the substrates for RebC and StaC appear to be the same, the source of differential products must lie with the use of a disparate FAD-based mechanism, achieved by the manner in which each enzyme binds its substrate and the amino acid environment near the active site. Due to the high homology between these enzymes, small variations must be all that are required for this shift in mechanism. Here our structures of substrate-bound forms of RebC and the StaC-like catalyst RebC-10x are most informative. While our data show a common substrate, each enzyme active site is designed to bind a different tautomer of that substrate. A nonplanar keto tautomer of 7-carboxy-K252c is bound to the StaC-like catalyst while RebC binds the planar enol tautomer. Notably, there is a key difference in the chemical reactivity of these two tautomers: in the keto tautomer, the sp^3 hybridized C-7 carbon can accept electrons from the spontaneous decarboxylation of 7-carboxy-K252c (Figure II.11.A). In the enol form,

33

however, spontaneous decarboxylation is not possible due to the sp^2 hybridized C-7 carbon (Figure II.11.B). Thus, RebC binds 7-carboxy-K252c such that decarboxylation would require a change in C-7 hybridization, likely achieved through hydroxylation of C-7 in a FAD-dependent monooxygenase reaction (13, 25) (Figure II.11.C), while the StaC-like catalyst binds 7-carboxy-K252c primed for decarboxylation without the need for hydroxylation (Figure II.11.A). Support for an FAD-dependent monooxygenase mechanism for RebC includes the observation that its 'mobile' flavin adopts an 'in' position in the presence of substrate or when the FAD is reduced, such that the substrate C-7 position is perfectly positioned near the C4a of the isoalloxazine ring for attack on C4a-hydroperoxy-FAD, a common flavin-dependent monooxygenase intermediate (13, 26).

While there is an obvious role for FAD in the RebC mechanistic proposal, its role in a StaC-like catalyst is less clear. As described above, our data and that of the Onaka laboratory support StaC as a FAD-dependent enzyme (17), and while our redox measurements show that FAD potentials are virtually unchanged among RebC and StaC-like catalysts in the absence of substrate, our structural data do not buttress the idea that StaC is a monooxygenase. Instead, our data suggest that FAD could contribute to the catalysis of decarboxylation through a steric and/or electrostatic mechanism. In an overlay of the substrate-bound structures of RebC and RebC-10x, the carboxyl group of substrate in RebC-10x clashes with the FAD isoalloxazine moiety (Figure II.12.B-C), indicating that FAD binding and/or movement of the cofactor from the 'out' to the 'in' position could play a steric role in driving decarboxylation. Alternatively, or in addition, reduced FAD swinging 'in' to the active site could serve to facilitate decarboxylation through an electrostatic mechanism. Structural data presented here show that like RebC, the RebC-10x FAD does transition from 'out' to 'in' positions. In particular, we observe FAD 'out' in the native RebC-10x structure and find FAD 'in' with the product K252c bound. In contrast to RebC, where the FAD is 'in', positioned near substrate following the week-long CPA soak (13), here we find that FAD dissociates from RebC-10x crystals during the same time period. While in agreement with a weaker FAD binding affinity for RebC-10x, FAD dissociation during the soaking experiment is also consistent with the notion of an unfavorably crowded active site following substrate binding.

These two different mechanistic hypotheses for RebC and StaC-like catalysts are also in alignment with the observed FAD binding affinities. For RebC, tight binding of FAD is likely important for the intricate control of the FAD cofactor required for substrate hydroxylation, whereas for StaC-like catalysts, precise control over FAD positioning would not be required in a steric or electrostatic mechanism. Although tighter binding of FAD from micromolar to high

nanomolar appears to make for a better StaC catalyst, in vivo, weaker binding of the cofactor could allow for a single FAD to service multiple StaC enzymes. In its babysitter role, StaC protects the 7-carboxy-K252c from unwanted side-reactions while waiting for FAD to bind. While organisms containing the rebeccamycin pathway could also benefit from shared use of FAD by multiple RebC enzymes, in this case, weaker FAD affinity might impair catalysis to too great an extent. Here we find that weakening FAD affinity from low to high nanomolar leads to a poor RebC catalyst.

While the binding of different tautomers of 7-carboxy-K252c by RebC and StaC-like catalysts can explain how two highly homologous FAD-dependent enzymes can use different mechanisms to yield their observed varied products, it is also interesting to consider the number of amino acid changes that are required for this differential binding of substrate. Based on structural analysis, two arginines (R239 and R230) in RebC are responsible for interacting with the 7-carboxyl moiety of 7-carboxy-K252c and appear key to maintaining it in an enol form, requiring RebC-like chemistry for product formation (see Figure II.10.D and Results). In StaC-like enzymes, the equivalent of R239 is asparagine and R230 adopts an altered position due to G48S and F216V substitutions. As we were preparing this manuscript, two groups reported that conversion of RebC activity to StaC-like activity can be achieved via two simultaneous mutations, F216V and R239N (17, 18), with individual mutations of these residues leading to unchanged and dead enzyme, respectively (17). These mutational data highlight the importance of the residue at position 239 and show that the availability of the 'lower' (StaC-like) conformation of R230, afforded by the F216V mutation alone, is not enough to convert a RebC to a StaC. The authors also tried the complementary mutations in StaC and found that this construct, StaC-V221F-N244R, is not an active RebC catalyst (17). Consistent with this result, our structural data suggest the StaC R230 equivalent, R235, cannot adopt its "RebC-like" orientation unless S48 is converted to G (see Figure II.10.D).

Collectively, these data provide insight into the key amino acid interactions in the RebC/StaC active sites that afford the product specificity. For RebC activity, R239 and an 'upper' position of R230 appear to help stabilize an enol form of 7-carboxy-K252c, requiring the tightly bound FAD to perform monooxygenase activity for product formation. In StaC, the absence of arginine at position 239 and a 'lower' position of R230, appear to yield the preferential binding of the keto form of 7-carboxy-K252c. Chemical logic predicts that this keto tautomer would not require FAD monooxygenase activity for conversion, perhaps using the 'mobile flavin' as steric or electrostatic driving force for decarboxylation instead. Of course after decarboxylation, protonation of the C-7 carbon is required to form the StaC product, K252c

(Figure II.11.A). Here we find that the 'lower' position of R230 creates an ordered water site that is not present in RebC (Figure II.10.E). This water is in position to act as a general acid to protonate C-7, yielding the StaC product. Interestingly, this water molecule is present regardless of the observed substrate orientation. As such, both orientations of 7-carboxy-K252c shown in Figure II.10.B–C seem to be catalytically viable. In contrast, only one orientation of substrate is found in the RebC structures and only that orientation has C-7 positioned correctly for hydroxylation.

II.V   MATERIALS AND METHODS

Generation of expression vectors for wild-type proteins

RebC was cloned into pET28a (Novagen) as previously described (12). *StaC* was amplified via PCR from *Streptomyces longisporoflavus* genomic DNA using primers 5'-*ggagag***CATATG**ACGCATTCCGGTGAGCGGAC-3' and 5'-*gtc***AAGCTT**TCAGCCCCGCGGCTCACGGGGCG-3' (Integrated DNA Technologies), where italicized text indicates an overhang and bold text indicates a restriction digest site. The PCR reaction mixture contained 1.25 U of *Pfu* Turbo DNA polymerase (Stratagene), 1x cloned Pfu buffer (Stratagene, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 µg/mL bovine serum albumin), 5% DMSO, 250 µM dNTPs (an equimolar mixture dATP, dCTP, dGTP, and dTTP), 1 µL of purified genomic DNA, and 0.5 mM of each primer. The 100 µL reaction mixtures were subjected to the following PCR cycle: 94 °C (3 min); 30 cycles of 94 °C (1 min), 52, 56, 60 or 63 °C (1 min), and 72 °C (2.5 min); 72 °C (10 min). Annealing temperatures of 56, 60 and 63°C all gave successful PCR products, so these were all combined. The PCR fragment was gel purified using the QIAquick Gel Extraction Kit (Qiagen) and digested at 37°C with *NdeI* and *HindIII* (New England Biolabs) in NEBuffer 2 (New England Biolabs, 10 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9). Purified pET28a (Novagen) was identically digested with *NdeI* and *HindIII* (New England Biolabs). Both digested PCR product and pET28a were again gel purified. The *staC* insert was ligated into the cut pET28a vector using T4 DNA Ligase (New England Biolabs, 2000 U) in 1x T4 DNA Ligase buffer (New England Biolabs, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP), with incubation at 16 °C for 12 h. Ligation reactions were used to transform TOP10 chemically competent *E. coli* cells, which were then plated onto LB-agar-kanamycin plates. Resulting colonies were amplified in LB-kanamycin, and the plasmid purified (QIAprep Spin

Miniprep Kit) then analyzed by restriction enzyme digestion (*NdeI/BglI*) for presence of the insert. The sequence integrity was confirmed by DNA sequencing (Molecular Biology Core Facility, Dana-Farber Cancer Institute, Boston).

InkE was amplified via PCR from *Nonomuraea longicatena* (15) genomic DNA using primers 5'-*ggagag***CATATG**ACTCGCAGCGAAGAGACCGAC-3' and 5'-*ccg***AAGCTT**TCACCCCGCCCCTCGCACGAGATC-3' (Integrated DNA Technologies), where italicized text indicates an overhang and bold text indicates a restriction digest site. The PCR reaction mixture contained 1.25 U of *Pfu* Turbo DNA polymerase (Stratagene), 1x cloned Pfu buffer (Stratagene, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 µg/mL bovine serum albumin), 5% DMSO, 250 µM dNTPs (an equimolar mixture dATP, dCTP, dGTP, and dTTP), 0.25 µL of purified genomic DNA, and 0.5 mM of each primer. The 50 µL reaction mixtures were subjected to the following PCR cycle: 98 °C (3 min); 30 cycles of 98 °C (1 min), 52, 59, 66 or 72 °C (1 min), and 72 °C (2.5 min); 72 °C (10 min). All annealing temperatures (52, 59, 66 and 72 °C) gave successful PCR products, so were all combined. The PCR fragment was gel purified using the QIAquick Gel Extraction Kit (Qiagen) and digested at 37°C with *NdeI* and *HindIII* (New England Biolabs) in NEBuffer 2 (New England Biolabs, 10 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9). Purified pET28a (Novagen) was identically digested with *NdeI* and *HindIII* (New England Biolabs). Both digested PCR product and pET28a were again gel purified; the *inkE* fragment was then concentrated five-fold by lyophilization and re-suspended in water. The *inkE* insert was ligated into the cut pET28a vector using Quick Ligase (New England Biolabs, 2000 U) in 1x Quick Ligase buffer (New England Biolabs, 66 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM ATP, 7.5% polyethylene glycol (PEG 6000)), with incubation at 25 °C for 3 h. Ligation reactions were used to transform TOP10 chemically competent *E. coli* cells, which were then plated onto LB-agar-kanamycin plates. Resulting colonies were amplified in LB-kanamycin, and the plasmid purified (QIAprep Spin Miniprep Kit) then analyzed by restriction enzyme digestion (*NdeI/HindIII*) for presence of the insert. The sequence integrity was confirmed by DNA sequencing (Molecular Biology Core Facility, Dana-Farber Cancer Institute, Boston).

AtmC was amplified from *Actinomadura melliaura* (16) genomic DNA (a kind gift of Carl Balibar) via PCR. A reaction mixture with 2.5 U of *Pfu* Turbo DNA Polymerase (Stratagene), 1x cloned *Pfu* buffer (Stratagene, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 µg/mL bovine serum albumin), 5% DMSO, 300 µM dNTPs

(an equimolar mixture dATP, dCTP, dGTP, and dTTP), 1 µL of purified genomic DNA, and 125 ng of each primer (5'-*caagtta*CATATGACCACGGCTTACGAGACCGA-3' and 5'-*caagtta*GAATTCCCATGAAGACCAGCCAGTTCTCCA-3', purchased from Integrated DNA Technologies) in a 50 µL reaction mixture was subjected to the following PCR cycle: 95°C (2 min); 30 cycles of 95°C (30 sec), 57°C (30 sec), and 72°C (3 min); and 72°C (10 min). The PCR fragment was gel purified using the QIAquick Gel Extraction Kit (Qiagen) and digested with *NdeI* and *EcoRI* (New England Biolabs) at 37°C in NEBuffer EcoRI (New England Biolabs, 50 mM NaCl, 100 mM Tris-HCl, 10 mM $MgCl_2$, 0.025% Triton X-100, pH 7.5). Purified pET28a (Novagen) was identically digested with *NdeI* and *EcoRI* (New England Biolabs). Both digested PCR product and pET28a were again gel purified. Various ratios of insert to vector were ligated in a 10 µL reaction volume using 1x T4 DNA ligase buffer (New England Biolabs, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, 25 µg/mL bovine serum albumin) and 400 U of T4 DNA ligase (New England Biolabs), with incubation at room temperature for 50 min. Ligation reactions were transformed into chemically competent DH5α cells and plated onto LB-agar-kanamycin plates. Resulting colonies were amplified in LB with 30 mg/L kanamycin, and the plasmid was purified (QIAprep Spin Miniprep Kit) and assayed via restriction digestion for incorporation of the insert. The integrity of the sequence was assayed with sequencing reactions spanning the length of the insert. Sequencing was carried out at the MIT Biopolymers Laboratory.

Generation of the StaC-10x expression vector

Twenty-eight overlapping primers encoding the N-terminal portion of the StaC protein with ten amino acid substitutions were designed using DNAWorks 2.4 (http://mcl1.ncifcrf.gov/dnaworks/dnaworks2.html) and synthesized by Integrated DNA Technologies. Primers were designed with each codon for a modified amino acid in a non-overlapping portion of a primer, such that each primer encoding a modified amino acid could be singularly 'swapped' with a primer encoding the wild-type amino acid (Figure II.13). This primer design scheme enables facile production of modified StaC proteins with fewer unwanted mutations.

Primer Number	Sequence[1]
1	5'-*ggagag*CATATGACCCACTCTGGCGAACGTACGGATGTT-3'
2	5'-TACCGACTGGACCACCACCAACGATTAAAACATCCGTACGTTCGC-3'
3	5'-GGTGGTCCAGTCGGTATGGCACTGGCGTTAGACTTACGCTACCGC-3'
4	5'-GACAACTAAACAATCAATGCCGCGGTAGCGTAAGTCTAACGCCAG-3'

5	5'-TGATTGTTTAGTTGTC<u>GAACAGACG</u>GATGGCACGGTCCGTCATCC-3'
6	5'-ATAGAACGCGGACCAATGGT<u>GCC</u>AAC<u>ACG</u>TGGATGACGGACCGTG-3'
7	5'-ATTGGTCCGCGTTCTATGGAGCTGTTCCGCCGTTGGGGTGCAGCA-3'
8	5'-CCGCTGGCCAACCGGCGTTACGGATTGCGTCTGCTGCACCCCAAC-3'
9	5'-GGTTGGCCAGCGGATCACCCATTAGATATTGCGTGGGTGACCAAG-3'
10	5'-ACGATAACGATAGATCTCATGACCGCCGACCTTGGTCACCCACGC-3'
11	5'-TCATGAGATCTATCGTTATCGTCGCGGTACGGCAGCGAATCGTCC-3'
12	5'-TGGTCTGGTTCCGGGGTGTGGACAAACGCTGGACGATTCGCTGCC-3'
13	5'-CCCGGAACCAGACCAAATCTGTCCGC<u>AAC</u>ACTGGTTAAACCCAGT-3'
14	5'-CTGGGTGCACGCCCACCGCCTCAATCAGCACTGGGTTTAACCAGT-3'
15	5'-GTGGGCGTGCACCCAGACGGCCCACTGTTATTATCTACGACCGTT-3'
16	5'-TCGTCGGTTTGGACGACGCCGTCAACGGTCGTAGATAATAACAGT-3'
17	5'-CGTCCAAACCGACGACCATGTCGAGGCGACGCTGACCGATCACGC-3'
18	5'-CGGGCACGCACGGTGCCGGTGGTACCGGTGGCGTGATCGGTCAGC-3'
19	5'-CGTGCGTGCCCGCTTTCTGGTTGCCTGCGACGGCGCAAGCTCTCC-3'
20	5'-TGGGGCTTCAATACCACAGGCACGACGAACCGGAGAGCTTGCGCC-3'
21	5'-GTGGTATTGAAGCCCCAGCACGCCATCGCACGCAGGTCTTCCGTA-3'
22	5'-TTTTAACTCCGGGGCACGGAATAAGATATTACGGAAGACCTGCGT-3'
23	5'-GTGCCCCGGAGTTAAAAGATCGCCTGGGTGAACGTGCGGCGCTG -3'
24	5'-AGCGTAACGTGCTGCTTAACATTAAG<u>AAA</u>TGAAACAGCGCCGCAC-3'
25	5'-GCAGCACGTTACGCTTTCCATTACGC<u>GCA</u>TTAAATGGTAGCGATT-3'
26	5'-ATCATCTGCACCAAC<u>CGT</u>CAGA<u>CGG</u>TATAAATCGCTACCATTTAA-3'
27	5'-GTTGGTGCAGATGATGATACGGGTGCCCGTCCGGATGCATTAGCA-3'
28	5'-*cgtcct***TGATCA**GTGCTAATGCATCCGGACGG-3'

[1]Bold text within oligonucleotide sequences indicates restriction digest sites, italicized text indicates modified overhangs, and underlined text indicates modified codons. Note that even numbered primers are reverse primers.

To generate the StaC-10x expression vector, a first PCR was carried out using 1x cloned *Pfu* buffer (Stratagene, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 µg/mL bovine serum albumin), 2.5 U of *Pfu* Turbo (Stratagene), 250 µM of dNTPs (an equimolar mixture of dATP, dCTP, dGTP, and dTTP), and 12.5 nM of each of the 28 primers, with a temperature cycle of 94°C (2 min); 45 cycles of 94°C (30 s), 58°C (30 s), and 72°C (1 min); and 72°C (2 min). A second PCR was then carried out using 1x cloned *Pfu* buffer, 2.5 U of *Pfu* Turbo, 250 µM dNTPs, 300 nM of each primer 1 and 28, and 1 µL of the first PCR, in a total volume of 50 µL using an identical temperature cycle. The product of the second PCR was gel purified using a QIAquick Gel Extraction Kit (Qiagen). Gel purified DNA from the second PCR was then incubated with 20 U of *Nde*I (New England Biolabs) in NEBuffer 3 (New England Biolabs, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.9) at 37°C for 2 h; then, after addition of 15 U of *Bcl*I (New England Biolabs), the reaction was incubated at 50°C for 2 h.

Separately, pET28a-StaC (12) was transformed into *dam⁻/dcm⁻ E. coli* competent cells (New England Biolabs) and purified using a minprep kit (Qiagen). Purified plasmid was incubated with 20 U of *Hind*III (New England Biolabs) in NEBuffer 2 (New England Biolabs, 10 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, 1mM DTT, pH 7.9) at 37°C for 2 h; then, after adjustment of the buffer to 50 mM Tris-HCl and 100 mM NaCl and addition of 15 U of *Bcl*I, incubated at 50°C for 2 h. Additionally, purified plasmid was separately incubated with 20 U of *Nde*I and 20 U of *Hind*III in NEBuffer 2 (New England Biolabs) at 37°C for 2 h, followed by heat inactivation of the restriction enzymes at 65°C for 20 min. 2.5 U of calf alkaline phosphatase (New England Biolabs) was then added to this final reaction mixture, which was incubated at 37°C for 2 h.

Appropriate fragments from the restriction digests were gel purified. Ligation reactions were set up with 20 ng of the purified *Nde*I/*Hind*III digest reaction, 8 ng of the purified *Hind*III/*Bcl*I reaction, 8 ng of the purified *Nde*I/*Bcl*I digest reaction, 1x T4 DNA ligase buffer (New England Biolabs, 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM ATP, 25 µg/mL bovine serum albumin), and 200 U of T4 DNA ligase (New England Biolabs) and incubated at 16°C for 24 h. Ligation reactions were transformed into DH5α cells and plated on LB-agar-kanamycin plates. Plasmid was purified from overnight cultures of single colonies, analyzed for the presence of the PCR insert via a diagnostic PCR reaction, and sequenced at the MIT Biopolymers Laboratory using three different primers to gather data from the entire length of StaC-10x gene.

Generation of the RebC-10x expression vector

Forty-one overlapping primers encoding the N-terminal portion of the RebC protein with ten amino acid substitutions were designed using DNAWorks 2.4 (http://mcl1.ncifcrf.gov/dnaworks/dnaworks2.html) and synthesized by Integrated DNA Technologies.

Primer Number	Sequence[1]
1	5'-*ggagag*CATATGAATGCCCC-3'
2	5'-CCAGGATCAGGACATCGGTTTCAATTGGGGCATTCATATGCTCTC-3'
3	5'-CGATGTCCTGATCCTGGGTGGCGGTCCAGTTGGTATGGCCTTAGC-3'
4	5'-CCCACCTGGCGATGGGCCAGGTCCAGGGCTAAGGCCATACCAACT-3'
5	5'-CATCGCCAGGTGGGTCACTTAGTCGTTGATGCGGGCGACGGCACC-3'
6	5'-TGGACCGATCGTAGAAACTTTTGGGTGCGTGATGGTGCCGTCGCC-3'
7	5'-GTTTCTACGATCGGTCCACGTAGCATGGAATTATTCCGTCGTTGG-3'
8	5'-GCCGTGCGAATCTGTTTTGCGACACCCCAACGACGGAATAATTCC-3'
9	5'-ACAGATTCGCACGGCGGGTTGGCCGGGTGACCACCCGTTAGACGC-3'

10	5'-TCATGGCCACCGACACGCGTCACCCATGCTGCGTCTAACGGGTGG-3'
11	5'-TGTCGGTGGCCATGAAGTTTACCGCATTCCATTAGGTACCGCGGA-3'
12	5'-CCGGGGTGTGTTCCGGCGTTGCACGCGTATCCGCGGTACCTAATG-3'
13	5'-GGAACACACCCCGGAACCAGATGCGATCTGTCCA<u>GCCG</u>CATTGGTT-3'
14	5'-TCGCCCACTGCCTCCGCCAGCAGCGGGGCTAACCAATG<u>CGC</u>TGGA-3'
15	5'-AGGCAGTGGGCGAGCGTTTACGTACGCGCTCTCGCTTAGACTCTT-3'
16	5'-CGCGCACGTGGTCGTCACGCTGTTCGAAAGAGTCTAAGCGAGAGC-3'
17	5'-GACCACGTGCGCGCCACGATTACGGATCTGCGTACGGGTGCAACC-3'
18	5'-GGCAACCAGGTAGCGCGCGTGAACTGCACGGGTTGCACCCGTACG-3'
19	5'-CGCTACCTGGTTGCCTGTGACGGCGCGTCTTCTCCGACCCGCAAG-3'
20	5'-GATGACGTGGTGGGGCGTCAATACCTAACGCCTTGCGGGTCGGAG-3'
21	5'-CCCCACCACGTCATCGCACGCAGGTCTTTCGTAATATTCTGTTCC-3'
22	5'-AAGCTGCGTAACTCCGGTGCACGGAACAGAATATTACGAAAGACC-3'
23	5'-CGGAGTTACGCAGCTTACTGGGCGAACGTGCGGCACTG<u>GTG</u>TTCT-3'
24	5'-GAAAACGTAAAGAAGAGCTCAGCATTAAGAAGAA<u>CAC</u>CAGTGCCG-3'
25	5'-TGAGCTCTTCTTTACGTTTTCCGTTACGC<u>TCT</u>CTGGATGGCCGTG-3'
26	5'-CGTCCACGCCCAC<u>AAC</u>TAA<u>GTTG</u>TATAAACCACGGCCATCCAGA<u>G</u>-3'
27	5'-GTGGGCGTGGACGATGCCTCTAAGTCTACCATGGACAGCTTTGAA-3'
28	5'-TCGAACGCCACTGCGCGACGAACTAATTCAAAGCTGTCCATGGTA-3'
29	5'-GCAGTGGCGTTCGATACGGAGATTGAAGTGTTATCTGATTCTGAA-3'
30	5'-ACGCGGTGCGTTAAGTGCCATTCAGAATCAGATAACACTTCAATC-3'
31	5'-CTTAACGCACCGCGTCGCGGATTCTTTCTCTGCGGGCCGTGTTTT-3'
32	5'-GCTTAAGGTATGTGCTGCATCGCCCGTCAGGAAAACACGGCCCGC-3'
33	5'-GCAGCACATACCTTAAGCCCAAGCGGTGGTTTCGGCATGAACACG-3'
34	5'-AACCCAGGTCGGCTGCAGAGCCAATGCCCGTGTTCATGCCGAAAC-3'
35	5'-AGCCGACCTGGGTTGGAAGTTAGCAGCGACGCTGCGCGGCTGGGC-3'
36	5'-TTCCTCCTCGTAGGTCGCTAATAAGCCCGGACCGGCCCAGCCGCG-3'
37	5'-CGACCTACGAGGAGGAACGTCGTCCGGTGGCGATTACCAGCCTGG-3'
38	5'-CGTACGGCGCAGGTTAACATTCGCCTCCTCCAGGCTGGTAATCGC-3'
39	5'-CCTGCGCCGTACGATGGATCGTGAGTTACCACCGGGTCTGCACGA-3'
40	5'-CCAAGAGGATCCCTCGCCGCGTGGGCCATCATCGTGCAGACCCGG-3'
41	5'-*ccaaga***GGATCCCTCGCC**-3'

[1]Bold text within oligonucleotide sequences indicates restriction digest sites, italicized text indicates modified overhangs, and underlined text indicates modified codons. Note that even numbered primers are reverse primers.

A first PCR was carried out using 1x cloned *Pfu* buffer (Stratagene, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 μg/mL bovine serum albumin), 2.5 U of *Pfu* Turbo (Stratagene), 250 μM of dNTPs (an equimolar mixture of dATP, dCTP, dGTP, and dTTP), and 12.5 nM of each of the 41 primers, with a temperature cycle of 94°C (2 min); 45 cycles of 94°C (30 s), 58°C (30 s), and 72°C (2 min); and 72°C (10 min). A second PCR was then carried out using 1x cloned *Pfu* buffer, 2.5 U of *Pfu* Turbo, 250 μM dNTPs, 7.5 μM of each primer 1 and 41, and 1 μL of the first PCR, in a total volume of 50 μL using an identical temperature cycle. The product of the second PCR was gel purified using a QIAquick Gel Extraction Kit (Qiagen). Gel purified DNA from the second PCR was then incubated with 20 U of *Nde*I (New England Biolabs) and 20 U of *Bam*HI (New England Biolabs) in NEBuffer 3 (New England Biolabs, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.9) and 0.1 mg/mL bovine serum albumin (New England Biolabs) at 37°C overnight. Restriction digest

reactions were then heat inactivated at 65°C for 20 min. Similarly, pET28a-RebC was incubated with 20 U of NdeI (New England Biolabs) and 20 U of BamHI (New England Biolabs) in NEBuffer 3 (New England Biolabs, 50 mM Tris-HCl, 10 mM MgCl$_2$, 100 mM NaCl, 1 mM DTT, pH 7.9) and 0.1 mg/mL bovine serum albumin (New England Biolabs) at 37°C overnight. Restriction digest reactions were then heat inactivated at 65°C for 20 min and then cooled on ice. 5 U of Antarctic Phosphatase (New England Biolabs) and 1x Antarctic Phosphatase Reaction Buffer (New England Biolabs, 50 mM Bis-Tris-Propane-HCl, 1 mM MgCl$_2$, 0.1 mM ZnCl$_2$, pH 6.0) were then added to this final reaction mixture, which was incubated at 37°C for 1.5 h. Phosphatase reactions were then heat inactivated at 65°C for 20 min.

Appropriate fragments from the restriction digests were gel purified. Ligation reactions with 20 ng of the purified NdeI/BamHI pET28a-RebC digest reaction and 27 ng of the NdeI/BamHI cut fragment encoding the N-terminus of RebC-10x were set up in 1x T4 DNA ligase buffer (New England Biolabs, 50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 10 mM dithiothreitol, 1 mM ATP, 25 µg/mL bovine serum albumin), and 200 U of T4 DNA ligase (New England Biolabs) and incubated at 16°C for 16 h. Ligation reactions were transformed into DH5α cells and plated on LB-agar-kanamycin plates. Plasmid was purified from overnight cultures of single colonies, analyzed for the presence of the PCR insert via a diagnostic PCR reaction, and sequenced at the MIT Biopolymers Laboratory using two primers to gather data from the entire length of Reb-10x gene. We observed from this sequence that our design of the fragment encoding the N-terminal fragment of RebC-10x inadvertently introduced a frame shift mutation. Therefore we used QuickChange (Stratagene) to introduce the missing cytosine, generating the final *RebC-10x* construct. Plasmid was again sequenced at the MIT Biopolymers Laboratory.

Protein purification

All protein expression vectors were transformed into chemically competent Rosetta™ 2 (DE3) pLysS cells (Novagen). Single colonies were used to inoculate LB media containing 30 mg/L kanamycin and 34 mg/L chloramphenicol. Cultures were grown at 37°C to an OD$_{600}$ of ~0.45, at which point the temperature was lowered to 21°C. At an OD$_{600}$ of ~0.65, cultures were induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for ~20 hours at 250 rpm. Cells were pelleted by centrifugation, suspended in 300 mM NaCl, 25 mM Tris-HCl, 5 mM imidazole, pH 8.0, sonicated, and centrifuged at 25,000 rpm to pellet insoluble material. The supernatant was incubated with nickel (II) loaded chelating sepharose fast flow (GE

Biosciences) for 40 min at 4°C. Unbound material was removed, and then the column was washed with 25 column volumes of 300 mM NaCl, 25 mM Tris-HCl, 20 mM imidazole, pH 8.0 and then, for AtmC and RebC, washed with 2 M KBr, 2 M urea, 300 mM NaCl, 25 mM Tris-HCl, 20 mM imidazole, pH 8.0 to remove bound flavin, and then re-washed with 300 mM NaCl, 25 mM Tris-HCl, 20 mM imidazole, pH 8.0. Protein was eluted with 300 mM NaCl, 25 mM Tris-HCl, 200 mM imidazole, pH 8.0. Eluted protein was loaded onto a Superdex-200 column pre-equilibrated with 150 mM NaCl, 25 mM HEPES, 10% glycerol, pH 7.5, and 3 mL fractions were collected. Fractions containing purified protein were pooled and used directly in isothermal titration calorimetry experiments. Protein concentration was assayed using absorbance in the linear range at A_{280}, using theoretical extinction coefficients calculated for each protein by ProtParam (27). For StaC-10x and RebC-10x, protein concentration was calculated instead using the BioRad Protein Assay (BioRad) calibrated using BSA (New England Biolabs) as a standard.

K_d for FAD of StaC, RebC, InkE, AtmC, StaC-10x, and RebC-10x measured by isothermal titration calorimetry (ITC)

All experiments were carried out using a MicroCal Isothermal Titration Calorimeter in the MIT Biophysical Instrumentation Laboratory. FAD was removed from RebC, AtmC, and RebC-10x as described above. Apo-protein from gel filtration was diluted in gel filtration buffer to the desired concentration. FAD could not be removed from StaC-10x in the same manner. After passing a buffer containing 2M urea and 2M KBr over a metal affinity column with bound StaC-10x to remove FAD, protein could not be eluted from the column using any tested elution buffer. Presumably, the protein aggregated or cross-linked to the column after the loss of FAD, requiring us to use StaC-10x with partial FAD occupancy in this experiments. The use of non-apo StaC-10x complicated the ITC analysis, since a fundamental assumption of Origin Software is that the ligand concentration in the cell is zero prior to the first injection of ligand. This limitation prompted a control experiment with RebC as described in text. The ITC experiments were repeated in at least triplicate, except for the StaC-10x experiment, which was done twice (Figure II.6). Equilibrium constants of binding (K_b's) were determined using the Origin Software and were converted to equilibrium constants of dissociation (K_d's). The single binding site model was used in determining K_b's.

Measurement of the redox potential of StaC-10x and RebC-10x

The midpoint potential of purified StaC-10x and RebC-10x were determined using a xanthine oxidase reduction method developed by Massey (20). Experiments were carried out at 25°C under anaerobic conditions in a glovebox using a S. II. Photonics 400 series spectrophotometer, housed in an MBraun Labmaster glove box. The cell solution was buffered by 25 mM HEPES (pH 7.0), and contained 10% glycerol, 150 mM NaCl, 1 μM benzyl viologen, 1 μM methyl viologen, 250 μM xanthine, 15 μM StaC-10x, and 50 μM 1-hydroxyphenazine (1-OHP) as a reference dye (-172 mV). Addition of a catalytic amount of xanthine oxidase (Sigma Aldrich) initiated the reaction, and full spectra were collected every 5 minutes until completion. The reduction of the StaC-10x flavin and 1-hydroxyphenazine were monitored at 480 nm and 370 nm, respectively. For this technique, unlike ITC, the level of flavin incorporation is thought to be irrelevant, as only the flavin cofactor itself is observed spectroscopically. Hence, barring interference of apo protein with the dye molecule (which is extremely unlikely), the presence of apo protein is not thought to influence the redox potential determined.

Activity assays of StaC, RebC, and StaC-10x

All enzyme assays were carried out as described earlier (12) in the presence of 75 mM HEPES pH 7.5, 1 mg/mL BSA, 5 mM NADPH, 20 μM ferredoxin, and 1 μM flavodoxin NADP$^+$ reductase. All assays were carried out in triplicate, except RebC, which was run once.

Crystallization and Structure Determinations

RebC-10x protein was purified as described above. After size exclusion, the protein sample was concentrated to 25 mg/mL (as determined by using the BioRad Protein Assay) using an Amicon Ultra 30K Centrifugal Filter Device in protein buffer (150 mM NaCl, 25 mM HEPES, 10% glycerol, pH 7.5), and flash frozen in liquid nitrogen for storage at -80°C. Crystals of RebC-10x were grown using the hanging-drop diffusion method at room temperature, by incubating 1 μL of protein sample (diluted to 20 mg/mL with protein buffer) with 1 μL of precipitant solution, 100 mM NH$_4$F and 20% PEG 3350 over a 0.5 mL well of precipitant solution. After 3-4 days, rod clusters would appear; these were used to generate microseeds using the Seed Bead kit (Hampton). The seed solution was diluted 100-fold in a 50% protein buffer, 50% precipitant solution and used to serially streak seed 3 drops of 1 μL of 4-6 mg/mL RebC-10x (diluted in protein buffer) and 1 μL of 50 mM NH$_4$F and 10% PEG 3350 which had

been pre-equilibrated for three hours. Single crystals appear after 1-2 days. Solutions of CPA and K252c for use in soaking experiments were synthesized as described (12) (Tocris Bioscience), and prepared at 50 and 10 mM in DMSO, respectively. The native/K252c structure was generated from crystals that were soaked in a solution of 25 mM NH_4F, 75 mM NaCl, 1 mM K252c, 12.5 mM HEPES (pH 7.5), 5% PEG 3350, and 5% glycerol for three hours at room temperature prior to cryocooling by washing the crystals through a soaking solution with 20% glycerol before submerging in liquid nitrogen. CPA soaked crystals were prepared by incubating single crystals in an identical soaking drop, but substituting K252c for 5 mM CPA. These crystals were incubated at room temperature for one week, then washed in cyroprotectant and cryocooled as described above.

Diffraction data to 2.15 and 2.37 Å resolution for K252c and CPA soaks, respectively, were collected at Beamline 24-IDE at the Advanced Photon Source (Argonne, IL), and processed with HKL2000 (28) to 2.33 and 2.76 Å resolution, respectively, trimming the resolution such that the I/s(I) is above 2.0 in the highest resolution bins (Table II.4). Both structures were phased using a previously solved RebC – CPA soaked structure (PDB ID 2R0G) stripped of water molecules and ligands as a model for rigid body refinement. Topology and parameter files for FAD, K252c, and S-keto-7-carboxy-K252c were adapted from those used previously (13). Iterative rounds of model building and refinement were done using the programs COOT (29), CNS (30, 31), and PHENIX (32). Non-crystallographic symmetry restraints were not used in either structure due to inherit differences of the molecules in the asymmetric unit. No sigma cutoff was included in the refinement and both structures were verified using composite omit electron density maps. The final R factors (work/free/all reflections) of 21.4/26.1/21.6 and 20.6/25.5/21.0 for K252c-soaked and CPA-soaked structures, respectively, are within the typically ranges for structures at or above 2.3 Å resolutions (33). In both structures, the 20-residue N-terminal tags containing the His_6 affinity sites are disordered. In addition, the RebC-10x K252c soaked structure is missing residues 1-3 and 247-252 in molecule A (K252c-bound structure) and residues 1-2, 353-364, 418-422 in molecule B (native structure) due to disorder. Water molecules were included at positions where +3.5σ F_o-F_c peaks were present. Water molecules in the RebC-10x CPA soaked structure were included only where positive F_o-F_c difference density was present and water binding sites were previously established by higher resolution structures. In this structure, residues 1, 40-41, 248-251, and 418-425 in molecule A and residues 1, 38-40, 247-252, and 418-425 in molecule B are not included in the model. See Table II.4 for refinement statistics.

II.VI ACKNOWLEDGEMENTS

Data deposition

The atomic coordinates and structure factor discussed in this chapter have been deposited in the Protein Data Bank (PDB ID codes 4EIP and 4EIQ).

Author contributions

K.S.R. helped design the study, generated *staC-10x* and *rebC-10x*, purified proteins except RebC-10x, carried out ITC experiments, performed initial activity assays, and contributed to writing of the paper. P.J.G. generated *rebC-10x*, purified RebC-10x, carried out ITC on RebC-10x, performed final activity assays, solved the crystal structures, and co-wrote the paper. M.J.H. determined the redox potential of StaC-10x and RebC-10x. A.R.H-J. generated the InkE, StaC, and RebC expression vectors and designed the previously reported activity assay. S.J.E. and C.T.W. provided mentorship. C.L.D. provided mentorship and co-wrote the paper.

Table II.1. Residues interchanged to generate the RebC-10x and StaC-10x constructs (see Figure II.2 for a complete sequence alignment).

RebC	StaC	Function
Glu[36]	Asp[37]	FAD binding (adenine)
Gln[37]	Ala[38]	FAD binding (adenine)
Thr[38]	Gly[39]	FAD binding (adenine)
Gln[117]	Ala[118]	FAD binding (phosphate)
Arg[46]	Lys[47]	FAD binding (isoalloxazine)
Gly[48]	Ser[49]	FAD binding (isoalloxazine)
Phe[216]	Val[221]	Substrate binding pocket
Ala[231]	Ser[236]	Substrate binding pocket
Arg[239]	Asn[244]	FAD (isoalloxazine) & substrate binding
Thr[241]	Val[246]	Substrate binding pocket

Table II.2. Dissociation constants for FAD of StaC- and RebC-like proteins, as determined by isothermal titration calorimetry (curves shown in Figure II.3).

Protein	# of ITC experiments	Dissociation Constant for FAD
StaC	3	15,000 ± 2,000 nM
InkE	4	59,000 ± 48,000 nM
RebC	5	20 ± 12 nM
AtmC	5	73 ± 28 nM
RebC-10x	3	600 ± 55 nM
StaC-10x-FAD[1]	2	560 ± 120 nM

[1]The FAD bound to StaC-10x could not be removed from the protein (see text).

Table II.3. Relative rates of arcyriaflavin A and K252c production (traces shown in Figure II.5).

Proteins assayed	Rate of arcyriaflavin A production (relative to StaP alone)	Rate of K252c production (relative to StaP alone)
StaP / StaC	1.1 ± 0.3	7 ± 1
StaP / RebC	14	1.0
StaP / RebC-10x	0.79 ± 0.07	20 ± 2
StaP / StaC-10x	1.7 ± 0.4	1.3 ± 0.2

Table II.4. Data collection and refinement statistics.

	RebC-10x K252c soak (K252c and native structures)	RebC-10X CPA soak (7-carboxy-K252c structure)
Data Collection		
Wavelength (Å)	0.97910	0.97918
Space Group	$P2_1$	$P2_1$
Cell Dimensions		
a, b, c (Å); β (°)	64.3, 78.6, 125.7; 99.9°	63.2, 77.7, 123.1; 98.8°
Resolution (Å)[1]	50 – 2.33 (2.42 – 2.33)	50 – 2.76 (2.86 – 2.76)
R_{sym} (%)[1]	8.2 (52.0)	5.4 (48.9)
$<I>$ / $\sigma(<I>)$[1]	19.7 (3.6)	19.6 (2.2)
Completeness (%)[1]	99.4 (98.5)	94.9 (97.4)
Redundancy[1]	5.2 (4.9)	3.5 (3.5)
Total reflections	256,655	101,523
Refinement		
Resolution (Å)	41.4 – 2.33	50 – 2.76
Reflections	51,610	28,820
R_{work} / R_{free}[2]	0.214 / 0.261	0.206 / 0.255
No. of non-hydrogen atoms		
Protein	7,872	7,812
FAD / Indolocarbazole	106 / 24	- / 54
Water	304	62
Average B-factors (Å^2)		
Protein	38.1	61.7
FAD / Indolocarbazole	40.2 / 31.1	- / 69.8
Water	35.3	52.3
Rms deviations		
Bond lengths (Å)	0.006	0.009
Bond angles (°)	1.0	1.6
Ramachandran statistics		
Most favored region (%)	88.5	85.7
Additionally allowed (%)	10.9	14.0
Generously allowed (%)	0.5	0
Disallowed (no. of residues)	1	2

[1]Highest resolution shell is shown in parentheses.
[2]R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of 5% of the diffraction data not used in refinement. The same test set of reflections used in the refinement of all structures.

Figure II.1. Indolocarbazole natural products and reaction schemes. (A) Chemical structures of rebeccamycin, AT2433-A1, staurosporine, and K252a with standard numbering. (B) Reaction scheme for StaP/StaC mediated production of K252c from CPA, and reaction scheme for StaP/RebC mediated production of arcyriaflavin A from CPA.

A.

Rebeccamycin

AT2433-A1

Staurosporine

K252a

B.

Chromopyrrolic acid (CPA)

StaP

Aryl-aryl coupled chromopyrrolic acid

OR

7-carboxy-K252c

StaC

RebC

K252c

Arcyriaflavin A

49
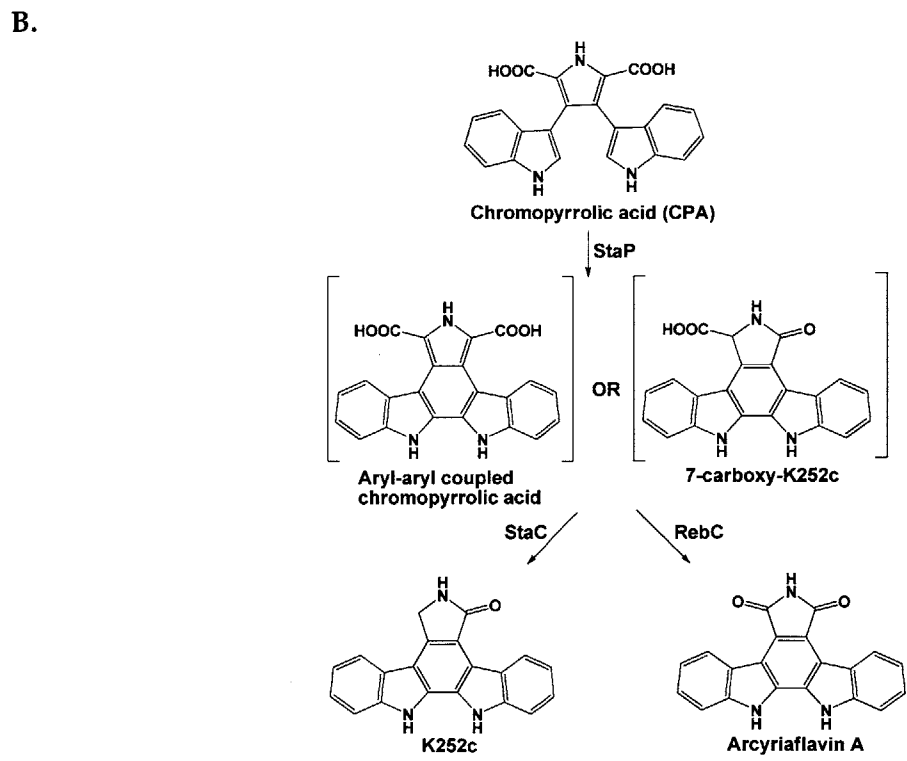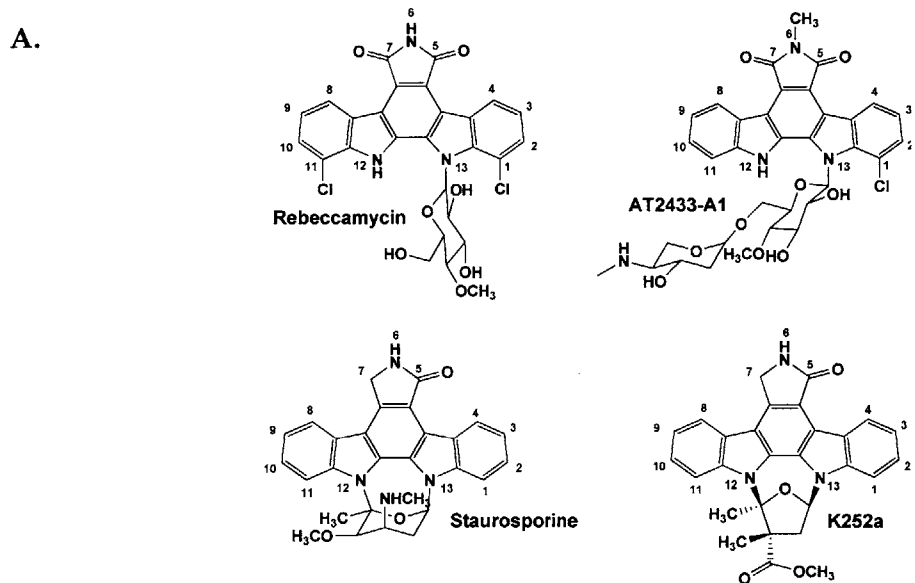
Figure II.2. ClustalW sequence alignment between the highly homologous enzymes AtmC, RebC, StaC, and InkE. High conservation is seen across all four proteins (percent identity between RebC and StaC, RebC and AtmC, StaC and InkE, and AtmC and InkE are 65, 64, 62, and 56, respectively). A star (*) represents residues conserved across all four proteins, and double dots or single dots represent similar (not identical) residues across all four proteins, according to the conventions of ClustalW. A black "x" above the alignment denotes residues involved in conserved FAD binding motifs common to flavin-dependent hydroxylases (12-14). To identify amino acids that may be responsible for the functional differences seen between the two pairs of enzymes (AtmC and RebC; StaC and InkE), the crystal structure of RebC in complex with FAD and 7-carboxy-K252c (PDB ID 2R0G) was analyzed. Residues where any atom is within 4 Å of FAD are highlighted in pink, residues where any atom is within 4 Å of putative substrate 7-carboxy-K252c are highlighted in turquoise, and residues where any atom is within 4 Å of both FAD and 7-carboxy-K252c are highlighted in yellow. Boxes indicate amino acids where the aligned residues are identical between StaC and InkE but distinct from the identical residues of RebC and AtmC. Note that some of these residues are not within 4 Å of either FAD or 7-carboxy-K252c. Green arrows point to mutation sites for the generation of RebC-10x and StaC-10x. RebC numbering for these residues are shown above the alignment.

```
                                xxxxxx                  36                46
                                                        378               48
AtmC  -MTTAYETDVLVIGGGPVGMALVLDLKYRGVGCQLIEASDGSVSHPRVGSIGPRSMELFR  59
RebC  -MNAPIETDVLILGGGPVGMALALDLAHRQVGHLVVEQTDGTITHPRVGTIGPRSMELFR  59
StaC  MTHSGERTDVLIVGGGPVGMALALDLRYRGIDCLVVDAGDGTVRHPKVSTIGPRSMELFR  60
InkE  -MTRSEETDVLIVGGGPVGLALSLDLTHRGVRHIVLDAGDGVVRHPKVSTVGPRSMEHFR  59
        .****::*****:** *** :*:  :    :::  **  :  ***:.::***** **

                                                                    117
AtmC  RWGIADRIRAAGWPGDHSLDTAWVTQVGGHEIHRLRVGTADTRPLPPYTPEPEQVCPQHW 119
RebC  RWGVAKQIRTAGWPGDHPLDAAWVTRVGGHEVYRIPLGTADTRATPEHTPEPDAICPQHW 119
StaC  RWGAADAIRNAGWPADHPLDIAWVTKVGGHEIYRYRRGTAANRPAFVHTPEPDQICPAHW 120
InkE  RWGVAGRVRDAGWPPGHPLDIAWVTRVGEYEIHRFERGTAAARPVFRHTPEPDQVCPAHL 119
        *** * :* **** .*.** ****:** :*::* *** *.  :****: :** *

                                                                xx
AtmC  LAPLLLEEARTHPGGVVRTRCRLDGFTQHDDHVEATVTDLAEGRELRIRARYMVAVDGAS 179
RebC  LAPLLAEAVGER----LRTRSRLDSFEQRDDHVRATITDLRTGATRAVHARYLVACDGAS 175
StaC  LNPVLIEAVGVHPDGPLLLSTTVDGVVQTDDHVEATLTDHATGTTGTVRARFLVACDGAS 180
InkE  LNPVLAGAVGRP----LRYLHRLEHFEQDAGCVRATIS--SNGEESVWARYLVACDGSS 173
        * *:*. .    :    ::  . *  . *.*:**:: *  :  : ::**::** **:*

                                                   216           231
AtmC  SPVRKACGIPSSARYDVMTFRNILFRAPELRARLGQR---EAMFYFLMLSNQLRFPVRAL 236
RebC  SPTRKALGIDAPPRHRTQVFRNILFRAPELRSLLGER---AALFFFLMLSSSLRFPLRAL 232
StaC  SPVRRACGIEAPARHRTQVFRNILFRAPELKDRLGER---AALVHFLMLSSTLRFPLRSL 237
InkE  SVVRKALGIDSPARHEAQVFRNVLFRAPGLPARLAERGHRSALYYLMHSWSLRYPMRSL 233
        * .*:**  :.  ::..*:   . .***:***** *   *.:* *:..:** **:*:*

      239
      241
AtmC  DGRSLYRLTVSGTDA-----DARDLVTMALAFETPVEILSDAVWHLTHRVAERFRQDRIF 291
RebC  DGRGLYRLTVGVDDASKSTMDSFELVRRAVAFDTEIEVLSDSEWHLTHRVADSFSAGRVF 292
StaC  NGSDLYNLVVGADDDTGARPDALALIKDALALDTPVELLGDSAWRLTHRVADFRYRAGRIF 297
InkE  DGRGLYNMVVDGRSDQ----GALELIRSAIAFDVPVELVADGLWHLTHRVADRYRAGRVF 289
        :*  .**.:**.:*.        .:   *: .  *:*::.*.  *:******: :   .*:*

          xx
AtmC  LLGDAAHTLSPSGGFGMNTGICAAADLGWKLAAELDGWAGRGLLDTYEEERRPVAVESLE 351
RebC  LTGDAAHTLSPSGGFGMNTGIGSAADLGWKLAATLRGWAGPGLLATYEEERRPVAITSLE 352
StaC  LAGDAAHTLSPSGGFGLNTGIGDAADLGWKLAATLDGWAGRHLLDTYDSERRPIAEESLN 357
InkE  LAGDAAHTLSPSGGFGMNTGIGDAADLGWKLAAVAGWAGDGLLGTYETERRPVAIAGLD 349
        * ******** ********:****  ********** : **** ** **: ****:*   .*:

AtmC  EANLNLRRTMGRPVPPELHLDTPAGAEARARMARQLALSDVAREFDAPGIHFGFTYRSSL 411
RebC  EANVNLRRTMDRELPPGLHDDGPRGERIRAAVAEKLERSGARREFDAPGIHFGHTYRSSI 412
StaC  EAHDNLRRTMKREVPPEIHLDGPEGERARAVMARRLENSGARREFDAPQIHFGLRYRSSA 417
InkE  AAEANLRRTVDRDLPAELAASTEKGARLRAEMAERLVNSGAREFDAPRVHFGFHYRSPI 409
        *. *****: * :*. :   .    *  ** :*:* *   ..***** :*** ***.

AtmC  IVAEPEQAPVDPRK----WQQSATPGARAPHAWLSPGASTLDLFGRGFTLLTFAEGAVGL 467
RebC  VCGEPE-TEVATGG----WRPSARPGARAPHAWLTPTTSTLDLFGRGFVLLSFGT----T 463
StaC  IVDDPD-VPVRQGQPDADWRPGSEPGYRAAHAWWDSTTSTLDLFGRGFVLLRFAD----H 472
InkE  VVSDGP---AEQGP---RWRPGSDPGCRAAHAWVRPGVSTLDLFGDGFTLLRFAD----S 459
        :  :      .          *:  .:  ** **.***   . .******* **.** *.

AtmC  EGVAGLERAFAERGVPLTTVRCDDRAVADLYEHPFVLVRPDGHVAWRAEAPPDDPGALAD 527
RebC  DGVEAVTRAFADRHVPLETVTCHAPEIHALYERAHVLVRPDGHVAWRGDHLPAELGGLVD 523
StaC  DGLPAIERAFAERGVPLTVHQGHDTEIAKLYARSFVLVRPDGHVAWRGDDLPGDPTALVD 532
InkE  PALSAFVTAFTERGVPFRSVLVGDPDTAALYGHRFVLVRPDGHVAWRGDDLEAAPANLAD 519
        .:  ..   **::* **:    ** : .:::.::.**.****.****.:  *.*

AtmC  LVRGGRR------ 534
RebC  KVRGAA------- 529
StaC  TVRGEAAPREPRG 545
InkE  LVRGAG------- 525
        ***
```

Figure II.3. Isothermal titration calorimetry data for FAD binding to StaC and homologues. ITC trace for (*A*). StaC; (*B*). RebC; (*C*). InkE; (*D*). AtmC; (*E*). RebC-10x; and (*F*). StaC-10x. Protein concentrations are: 29.6 µM (RebC), 30 µM (AtmC), 166 µM (StaC), 132 µM (InkE), 30 µM (StaC-10x), and 44 µM (RebC-10x). FAD was dissolved in the identical gel filtration buffer, and its concentration was calculated using A_{450} using $\varepsilon_{450} = 11,300$ cm^{-1} M^{-1}. FAD concentrations used in ITC experiments are: 300 µM (RebC experiment), 300 µM (AtmC experiment), 405 µM (StaC experiment), 1.05 mM (InkE experiment), 300 µM (StaC-10x experiment) and 300 µM (RebC-10x experiment). ITC experiments were carried out with protein in the cell and FAD in the syringe with the following parameters: 30 injections (one of 0.5 µL, 29 of 10 µL; with a duration time of 1 sec for the first injection and 20 sec for the remaining 29 injections), 240 sec spacing, 2 sec filter, 25°C, reference power of 15 µCal/sec, initial delay of 60 sec, stirring speed of 310, feedback set at high, ITC equilibrium set at fast and auto. For the InkE experiment, all conditions for ITC were identical, except that 60 injections were used (one of 0.5 µL, 59 of 5 µL; with a duration time of 1 sec for the first injection and 10 sec for the remaining 59 injections).

52

C.

D.

E.

F.

53
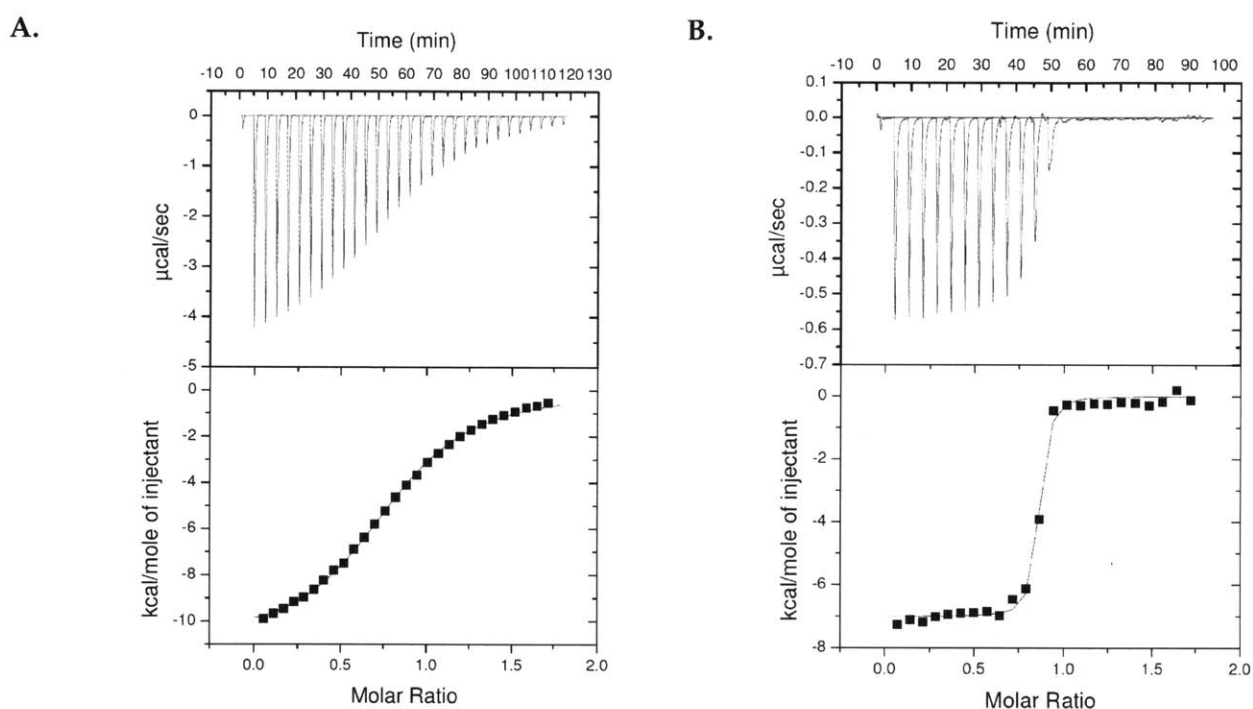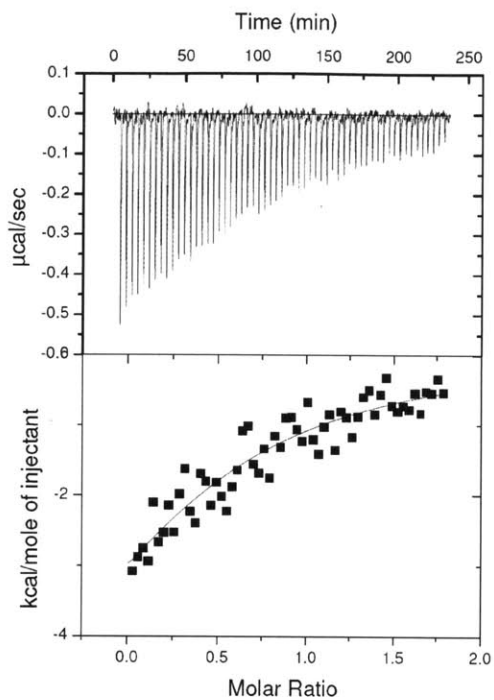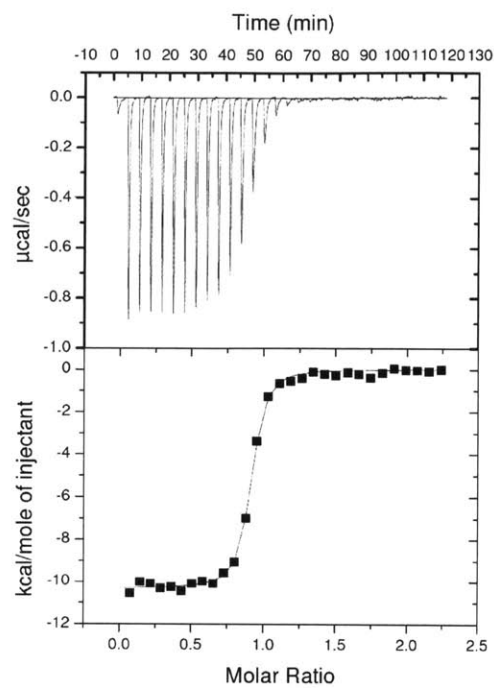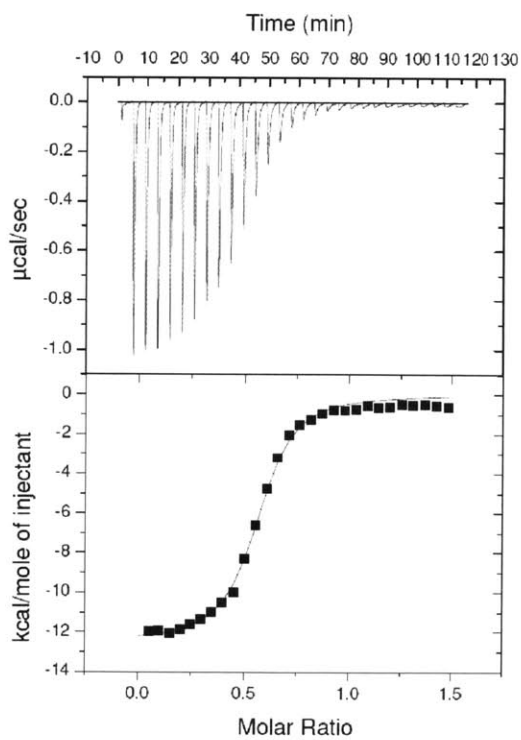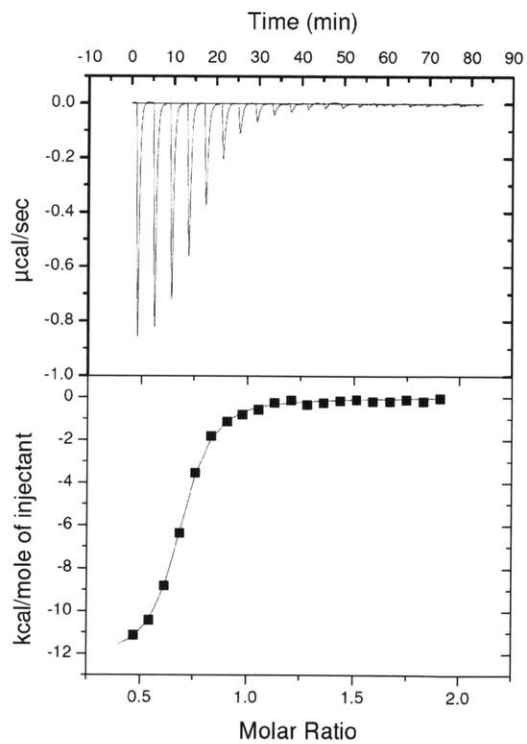
Figure II.4. Reduction potential of RebC-10x and StaC-10x. The redox potentials of (*A*). RebC-10x and (*B*). StaC-10x were determined using the xanthine/xanthine oxidase method (20). Inset: Plot of log([ox]/[red]) for 1-OHP vs log([ox]/[red]) for each enzyme, used to calculate the midpoint potential by comparing to 1-OHP (-172 mV).
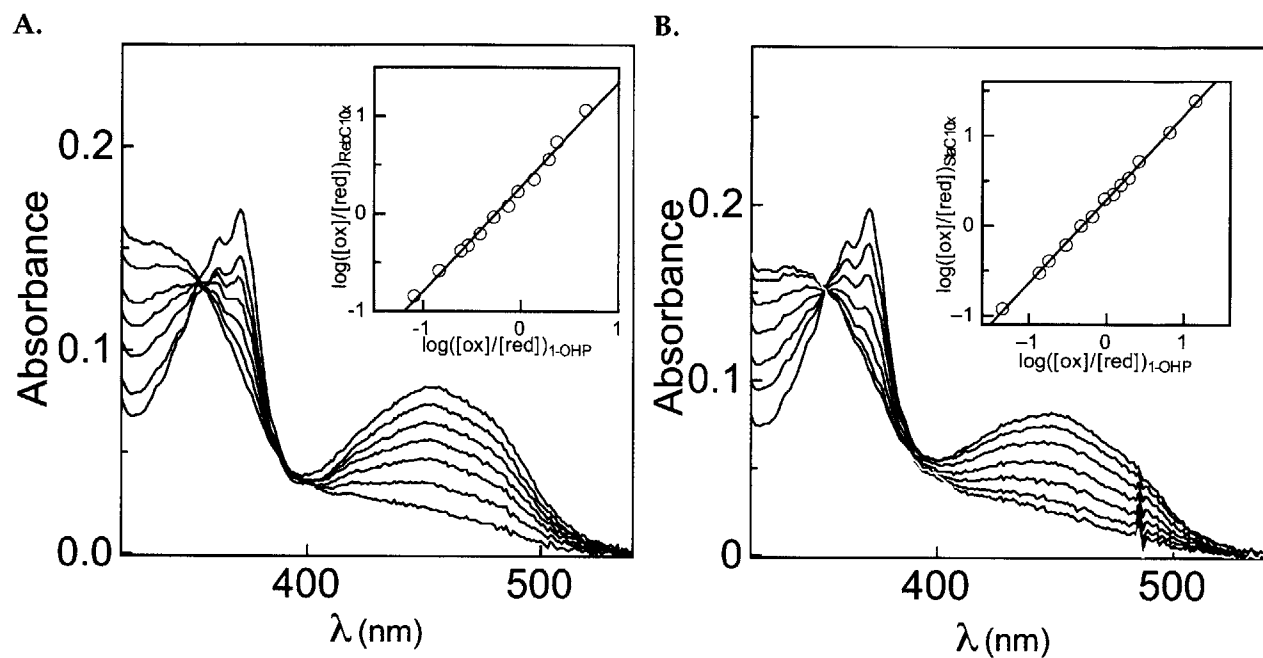
Figure II.5. Activity assay HPLC chromatograms. *(A).* 315 nm chromatogram. *(B).* 289 nm chromatogram. The retention time for arcyriaflavin A is 28.5 minutes. The arcyriaflavin peak in the RebC assay trace has been truncated. The retention time for K252c is 23.5 minutes. Peaks appearing at these times from experimental samples were integrated at the appropriate wavelength. Because reactions are linear at 30 min, integrated peak areas were converted to turnover rates using previously described methods (12). All enzyme assays were also carried out as described previously in the presence of 75 mM HEPES pH 7.5, 1 mg/mL BSA, 5 mM NADPH, 20 μM ferredoxin, and 1 μM flavodoxin $NADP^+$ reductase (34). RebC, StaC, StaC-10x, or RebC-10x were added at 5 μM. All reactions were initiated with the addition of 1 μM StaP and incubated at room temperature for 30 min. Reactions were then quenched with two volumes of methanol and incubated on ice for at least five minutes prior to the removal of precipitated protein by centrifugation. Reactions were assayed using an Agilent 1200 Series reverse-phase HPLC with an Agilent Eclipse C18 analytical column (150 x 4.6 mm). Two buffers (A is 0.2% trifluoroacetic acid in distilled, filtered water, and B is 0.2% trifluoroacetic acid in acetonitrile) were used in the following program of linear gradients at a 1 mL/min flow rate: 10 to 60% B over 30 min, 60 to 100% B over 0.5 min, 100% B for 2 min, 100 to 10% B over 0.5 min, and 10% B for 7 min.

A.
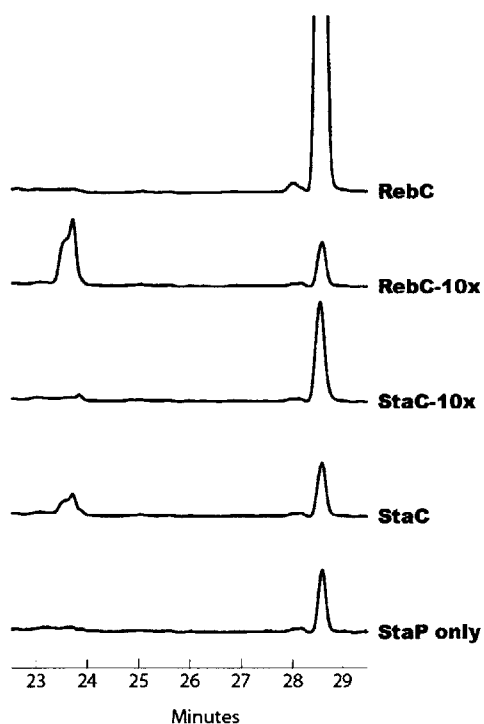
B.
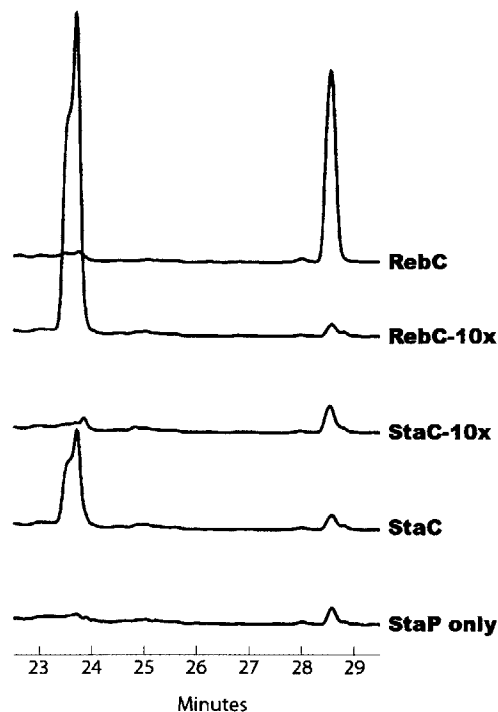
Figure II.6. Overall structure of RebC-10x aligned with wild type RebC. (A) RebC (lighter colors) and RebC-10x (darker colors) are aligned, showing the FAD binding domain (red), substrate binding domain (blue), thioredoxin-like domain (green), and FAD (grey sticks). RMSD of mainchain atoms is 0.46 Å. See Fig. S5 for comparison of (B) RebC-10x shows retention of the 'mobile flavin', where flavin binds in the 'out' position of native RebC-10x (cyan) and in the 'in' position when product is bound (slate), and the 'melting helix', where residues 354-363 (black) become disordered in absence of a bound indolocarbazole ligand (melting helix termini are marked with black circles in (A)). Arg230 is labeled to help relate this view of the active site with those shown in Fig. 3 and S6.
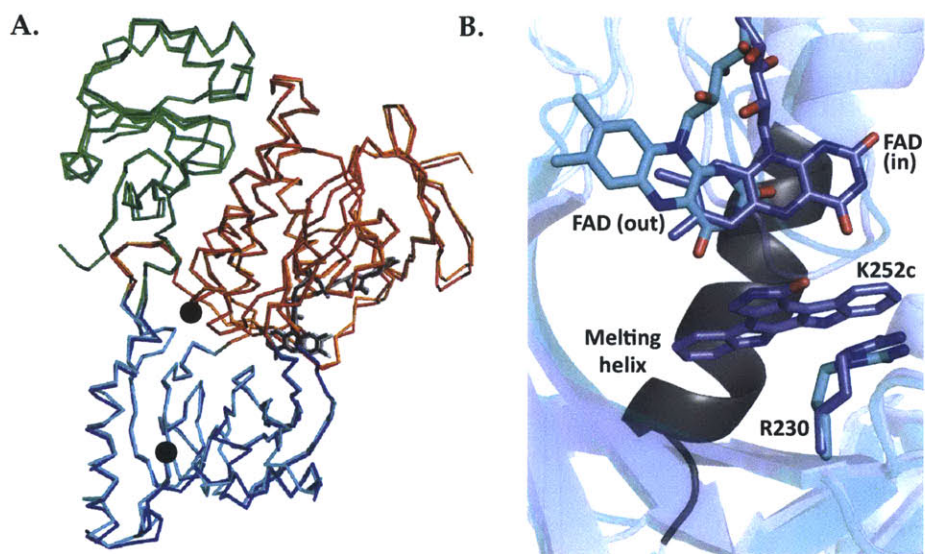
Figure II.7. B-factor analysis of RebC and RebC-10x in area of the ten mutations. (*A*). RebC (Native structure, PDB ID 2R0C) colored as a heat map from red (high b-factor) to blue (low b-factor). FAD is shown as grey sticks. (*B*). RebC-10x heat map based on "native structure", colored as above. (*C*). The locations of ten mutations in RebC-10x are colored red.
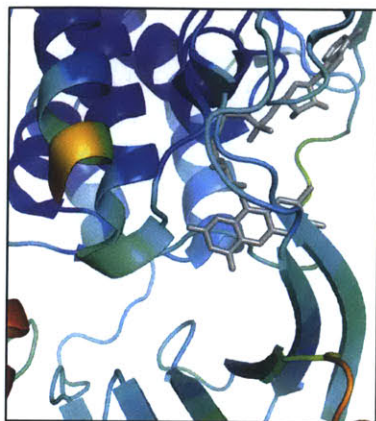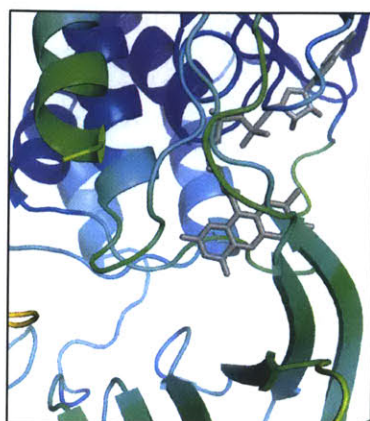
A.　　　　　　　　　　**B.**　　　　　　　　　　**C.**
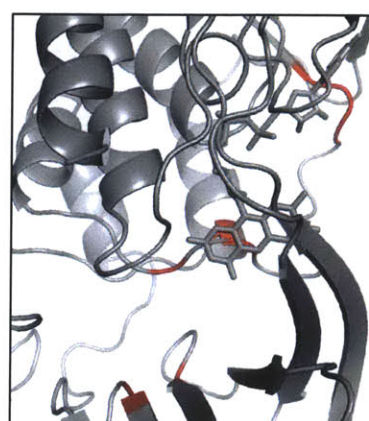
Figure II.8. Residues chosen from the RebC structures in the design of (i) StaC-10x and RebC-10x and (ii) subsequent RebC-10x residue positioning. Carbon coloring is as follows: native RebC backbone (yellow) and FAD (green), native RebC-10x backbone (cyan) and FAD (magenta), substrate-bound RebC (orange), and substrate-bound RebC-10x (gray). Glycines are shown in purple. *(A)*. A stretch of three residues from RebC – Glu[36], Gln[37], and Thr[38] – were installed in StaC-10x, replacing the StaC residues Asp[37], Ala[38], and Gly[39]. These residues may modulate binding of the adenosine portion of FAD. These residues are shown in the substrate-free structure of (i) RebC and (ii) RebC-10x, with their orientation unchanged in the structure with bound 7-carboxy-K252c. *(B)*. (i) Gln[117] in RebC interacts with a phosphate on the FAD via a water molecule. This residue was installed in StaC-10x, replacing Ala[118], which was installed in (ii) RebC-10x. Shown are Gln[117] and Ala[117] in the substrate-free structure of RebC and RebC-10x. The water binding site is preserved in RebC-10x. The orientation of both waters and residues are unchanged in the structures with bound 7-carboxy-K252c. *(C)*. (i) Arg[46] in RebC binds on the *si* face of the FAD in the substrate-free structure. Two residues away is Gly[48] (shown in purple). (ii) The corresponding residues in StaC are Lys[47] and Ser[49], and were installed in RebC-10x at positions 46 and 48. Also shown is Trp[276], which is unchanged in StaC, but is shown because it stacks on the *re* face of the FAD. *(D)*. As in *(C)*, but in the structures of RebC and RebC-10x are now bound with two different bound tautomers of 7-carboxy-K252c; the (i) enol and (ii) keto. (ii) Two different orientations of 7-carboxy-K252c were found in RebC-10x, one in each monomer; they are overlayed, one in magenta (molecule B) and one in thin grey lines (molecule A). (i) In RebC, Arg[46] now hydrogen bonds with a phosphate on the FAD. (ii) Flavin is not present in the RebC-10x substrate bound model and the loop containing Lys[46] has shifted downward. *(E)*. (i) Residues Phe[216], Ala[231], Arg[239], and Thr[241] in RebC line the active site pocket, however, only Arg[239] directly interacts with FAD, via a hydrogen bond with a carbonyl oxygen on the isoalloxazine ring. These residues were installed in StaC-10x, replacing Val[221], Ser[236], Asn[244], and Val[246], respectively. (ii) In RebC-10x, these StaC residues result in a more open binding pocket. Shown are the orientations of the (i) RebC and (ii) RebC-10x residues in the substrate-free structures. *(F)*. As in *(E)*, but in the structures of (i) RebC and (ii) RebC-10x with their respective bound tautomers of 7-carboxy-K252c. (i) Arg[239] now hydrogen bonds with the bound molecule in RebC; (ii) Asn[239] is too far to hydrogen bond to the molecule in RebC-10x.

D.

i)   R46   G48   W276   7-carboxy-K252c

ii)   S48   K46   W276   7-carboxy-K252c

E.

i)   F216   R239   T241   A231

ii)   V241   N239   V216   S231

F.

i)   7-carboxy-K252c   R239   F216   A231   T241

ii)   N239   V216   S231   T241
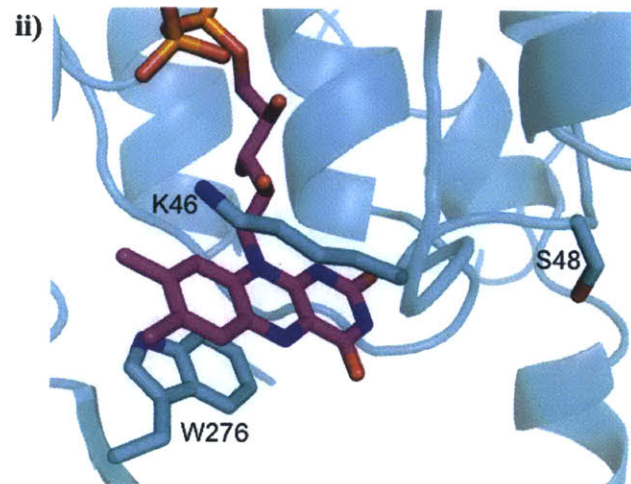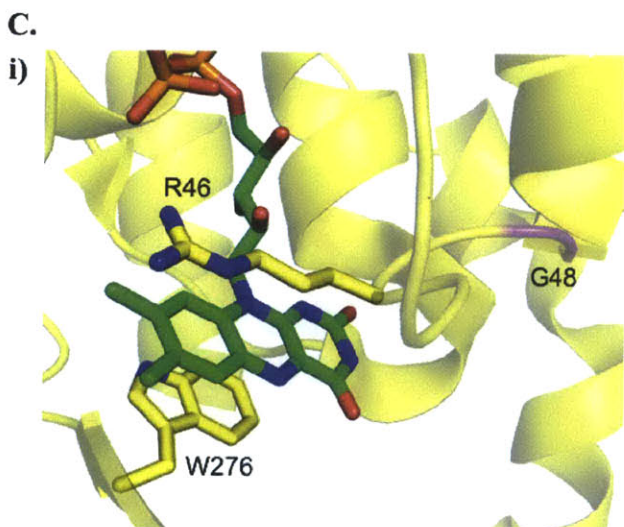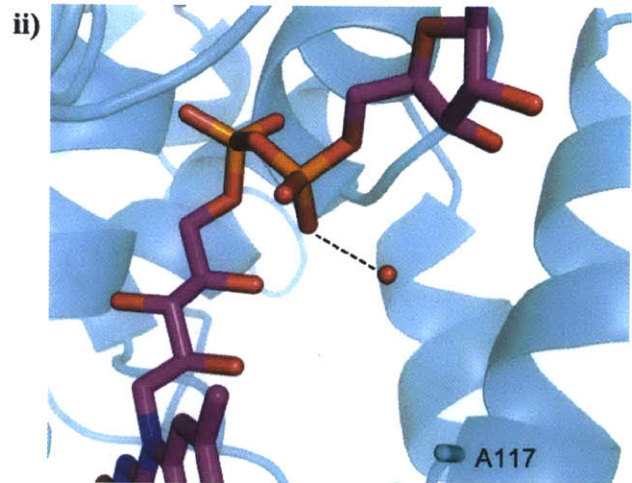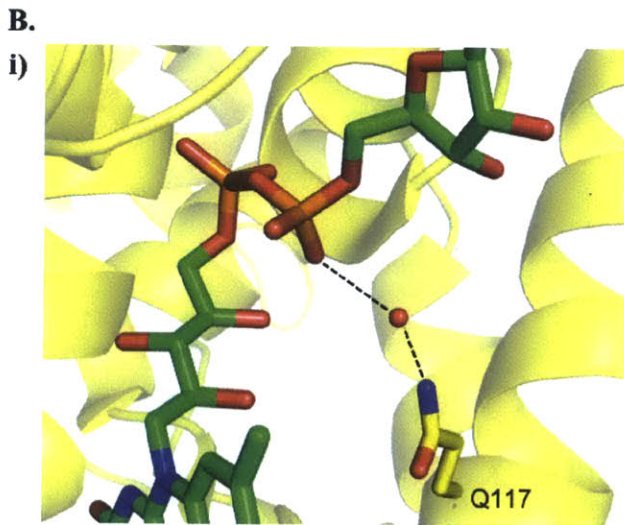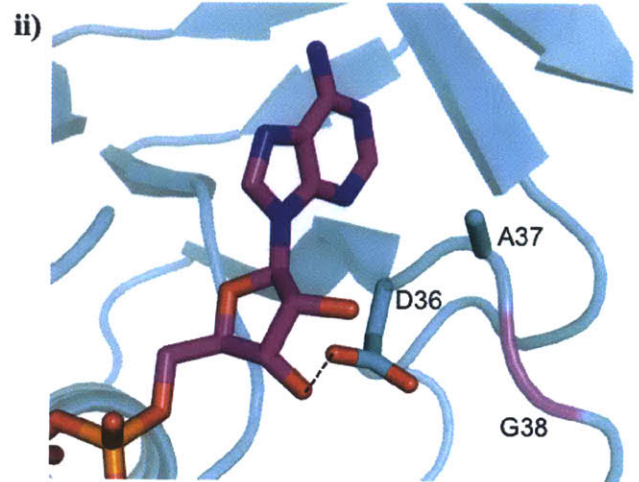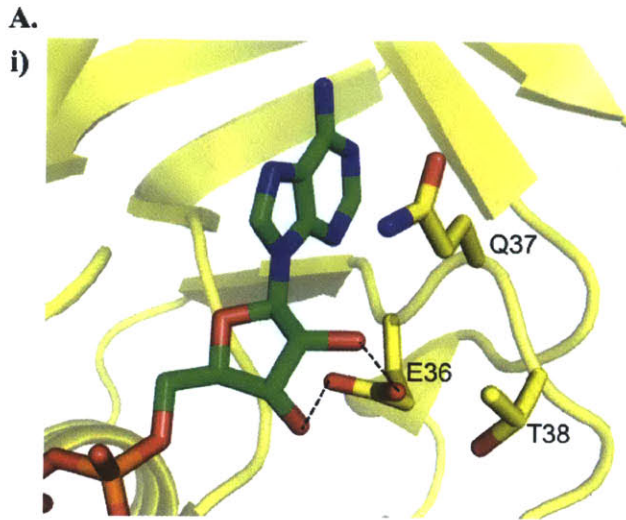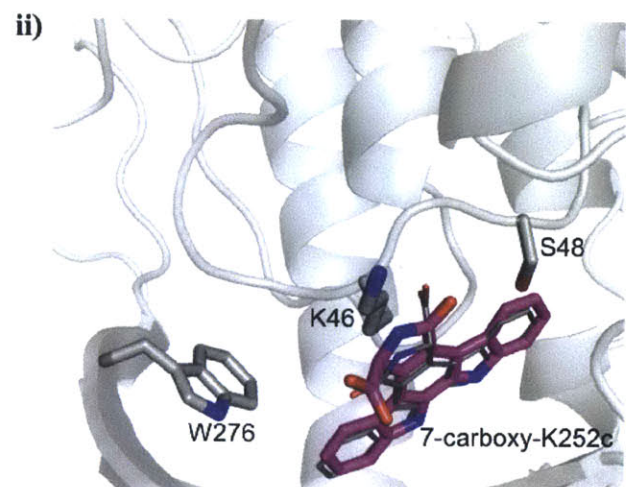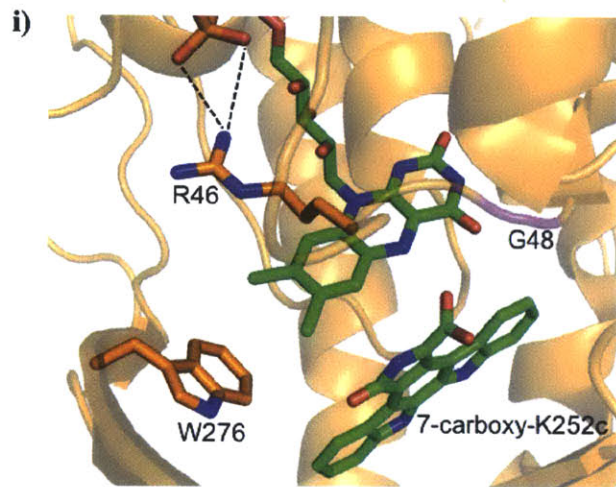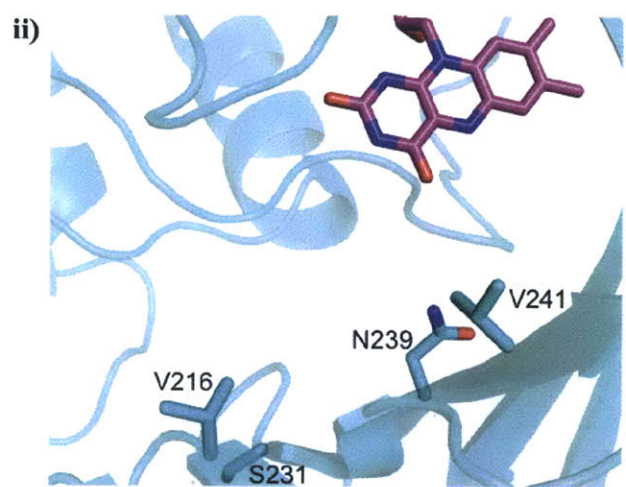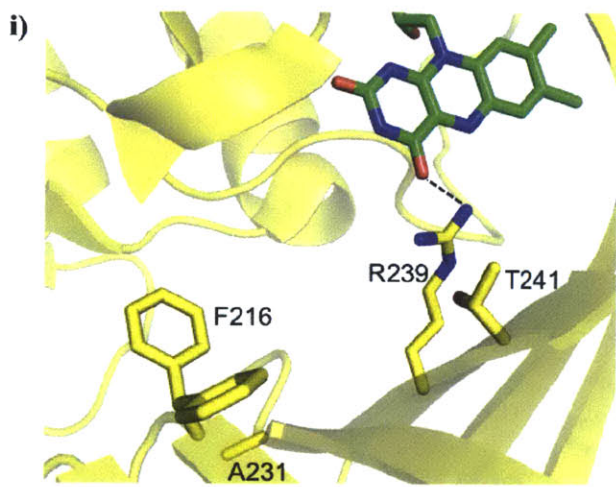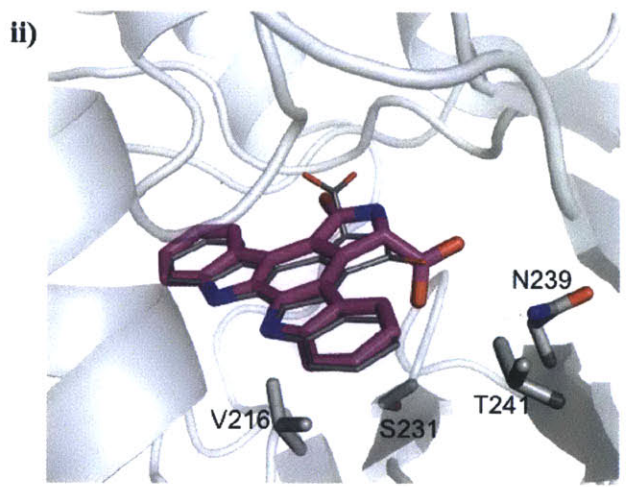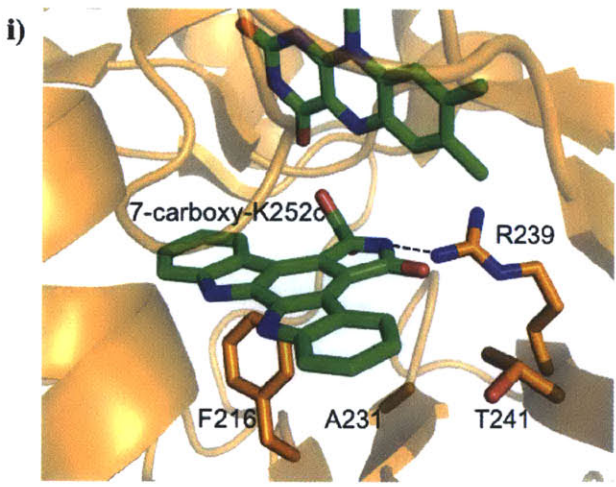
Figure II.9. Identification and orientation of bound indolocarbazoles in CPA-soaked RebC-10x. Difference electron density after refinement of the following molecules in the RebC-10x active site: (*A*). Arcyriaflavin A; (*B*). K252c; (*C*). Aryl-aryl coupled chromopyrrolic acid; (*D*). Enol tautomer of 7-carboxy-K252c; and (*E*). Keto tautomer of 7-carboxy-K252c. F_o-F_c difference electron density is contoured at +3.0σ (green) and -3.0σ (red), respectively. To show the relationship between this view and others, the box inlay shows the position of Arg230, a residue labeled in (*G*) and (*H*) below as well as in Figures II.6 and II.10. (*G*). One orientation of the bound keto tautomer of 7-carboxy-K252c with the carboxyl moiety pointed away from Arg230 is found in molecule B. (*H*). Another orientation of the bound keto tautomer of 7-carboxy-K252c with the carboxyl moiety pointed toward Arg230 is found in molecule A. Shown in pink mesh is a $2F_o$-F_c composite omit map contoured to 1.0σ.

61
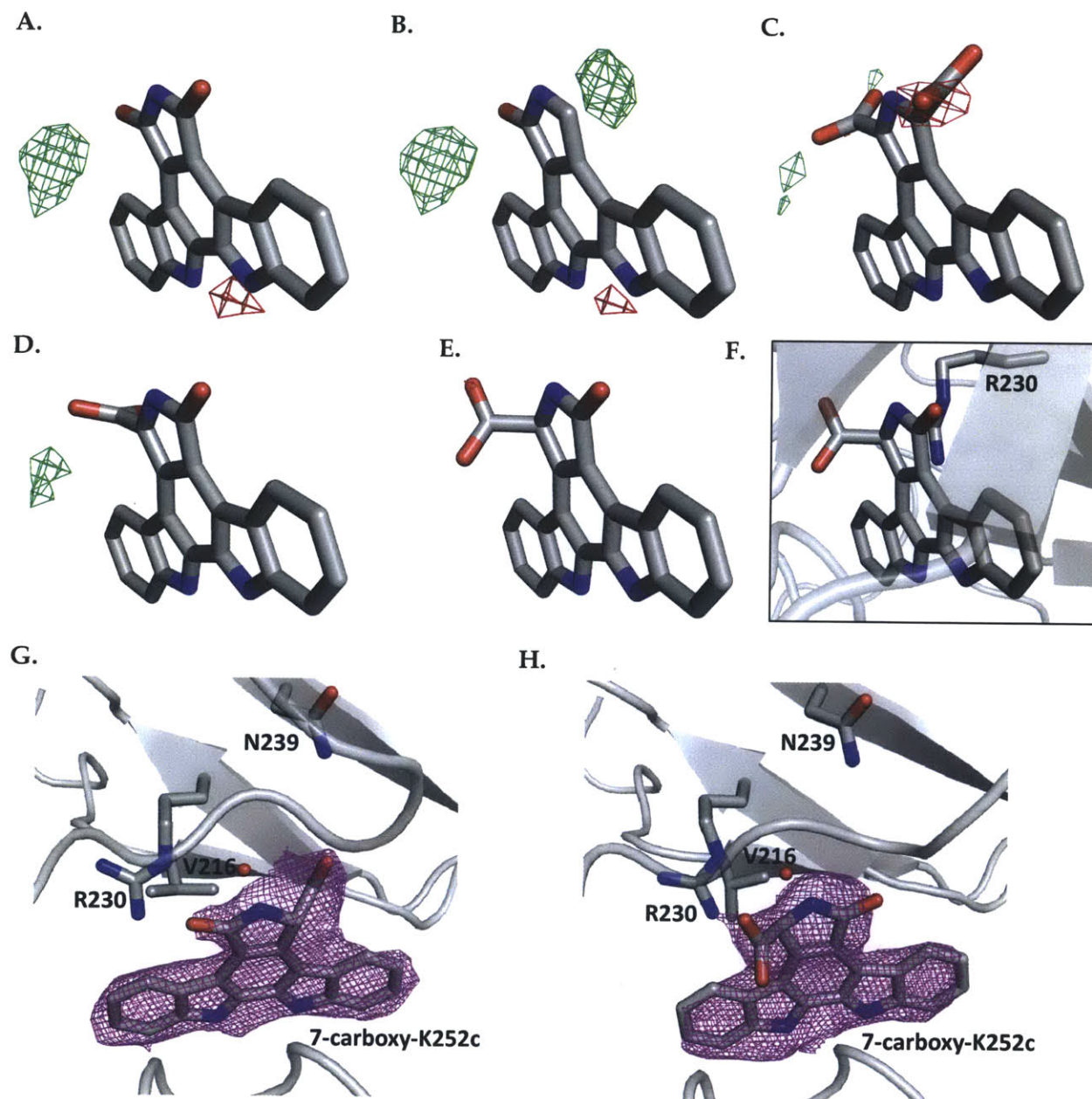
Figure II.10. RebC and RebC-10x substrate binding pocket. (A) RebC substrate binding pocket (PDB ID 2R0G) with protein backbone shown in orange and cofactor, FAD, and substrate, enol 7-carboxy-K252c, carbons in green and in stick representation. The sidechains, Arg230, Pro45, and Trp276 (representing a hydrophobic region of the binding pocket) are also shown in stick representation. (B) The *S*-keto form of 7-carboxy-K252c as found in molecule A of the RebC-10x CPA soaked crystal. Protein carbons are colored grey and substrate shown in stick representation and carbons colored magenta. The carboxyl group of substrate is proximal to Arg230, as in RebC. (C) The alternate orientation of substrate, as found in molecule B of the RebC-10x CPA soaked crystal. The carboxyl group of substrate is proximal to Trp276; coloring as in (B). (D) A stereo representation of an overlay of the RebC and RebC-10x (molecule A) active sites, shown from above. Substrate and protein colors the same as (A) – (C); RebC sidechains are labeled in parentheses. (E) A water molecule (red sphere) present in RebC-10x structure is located 2.6, 3.1 and 3.1 Å from two nitrogens of Arg230 and the C-7 position of 7-carboxy-K252c, respectively. In all stick representations, oxygen is shown in red and nitrogen in blue (see Figure S8A for the water binding site in the alternative orientation of substrate).
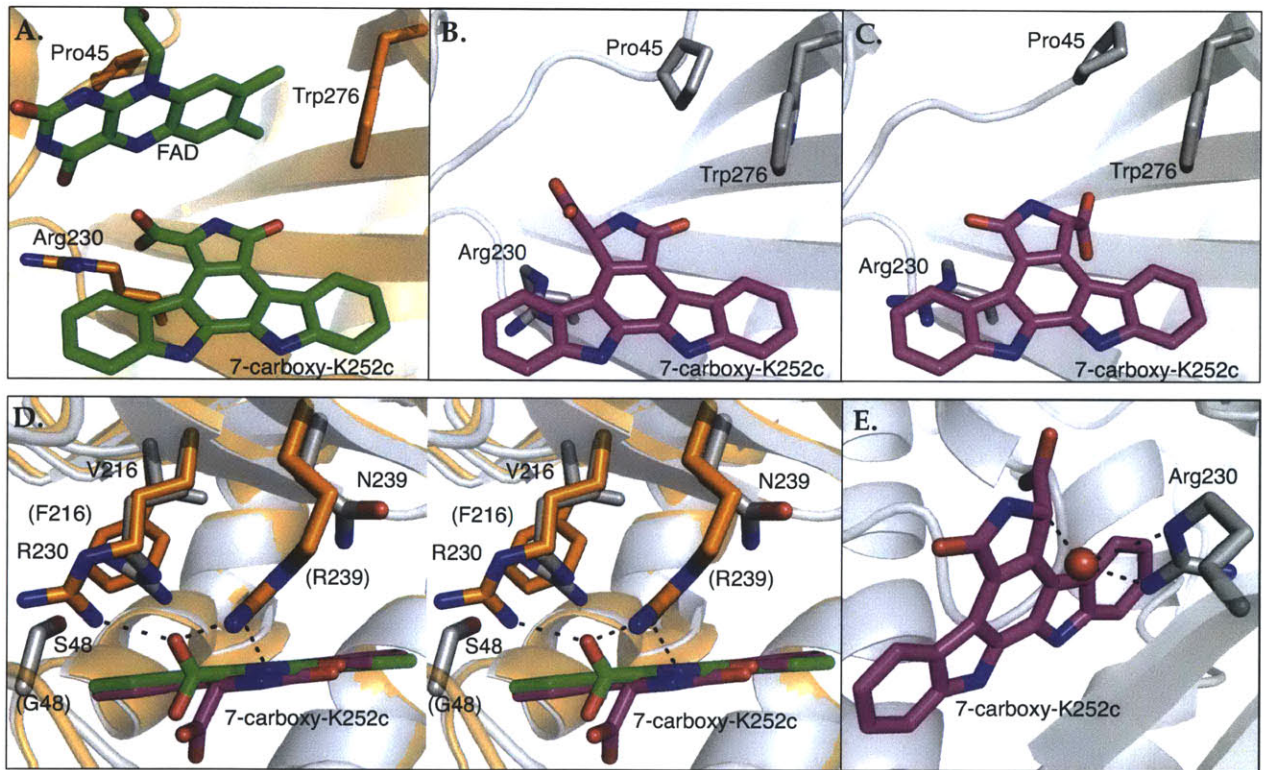
Figure II.11. Mechanistic implications of substrate tautomerization. (A) A mechanistic scheme for the decarboxylation of *S*-keto 7-carboxy-K252c in a StaC-like enzyme, where the reaction is initiated by decarboxylation. (B) Hybridization of the C-7 carbon in the enol tautomer of 7-carboxy-K252c makes decarboxylation as the first step in the mechanism unlikely. (C) A mechanistic scheme for RebC in which oxidation precedes decarboxylation at the C-7 position.

A.

7-carboxy-K252c (keto)

B.

7-carboxy-K252c (enol)

C.
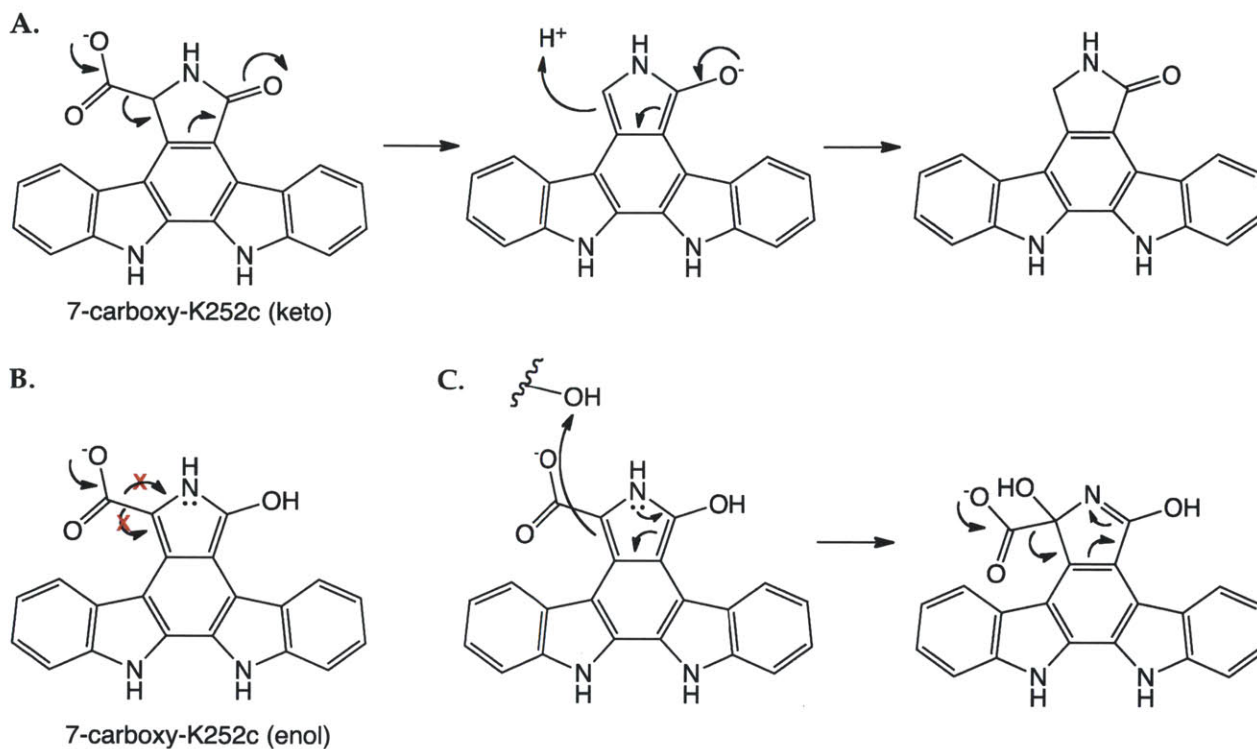
Figure II.12. Interactions of *S*-keto 7-carboxy-K252c with water and modeled flavin. *(A).* Water binding site for alternative orientation of substrate from that shown in Figure II.10E. The water is 2.9 and 3.4 Å from the Arg230 Nε position and the C-7 position of 7-carboxy-K252c, respectively. *(B).* Steric clash between FAD modeled in the "in" position and one orientation of bound *S*-keto 7-carboxy-K252c from the RebC-10x CPA soak structure. Colors: RebC-10x ribbons in grey, *S*-keto 7-carboxy-K252c carbons in magenta, RebC substrate-bound structure (PDB ID 2R0G) ribbons in orange, FAD and enol 7-carboxy-K252c carbons in green. *(C).* Steric clash between FAD modeled in the "in" position and the other orientation of bound *S*-keto 7-carboxy-K252c from the RebC-10x CPA soak. Colors as in *(B).*

A.

B.

C.
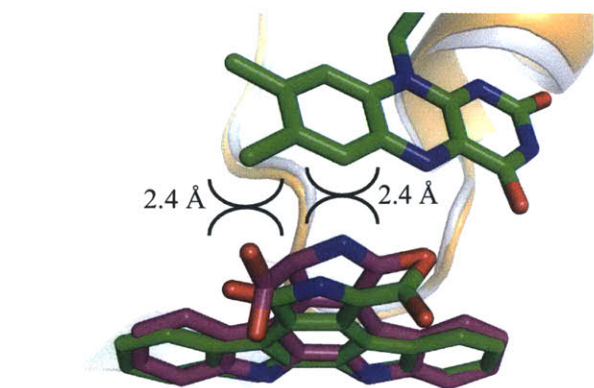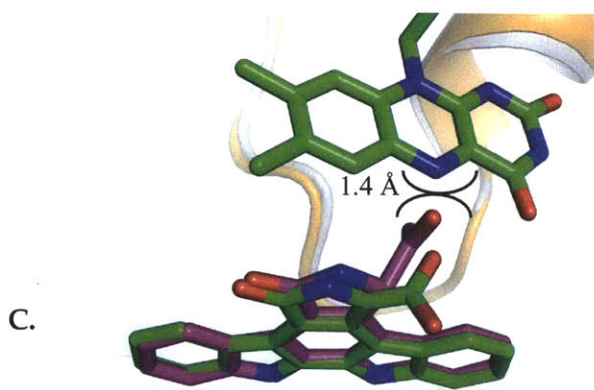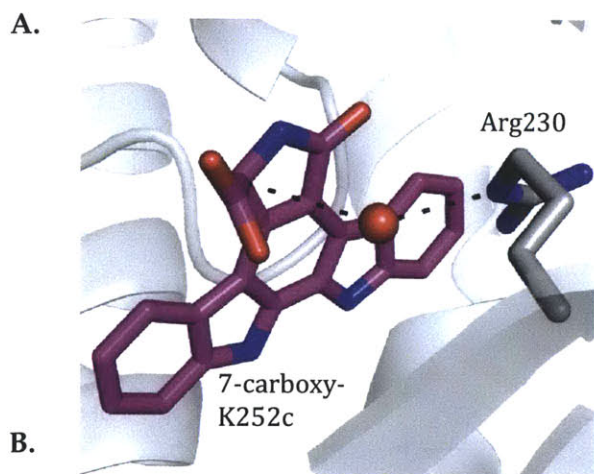
Figure II.13. Construction of the *staC-10x* gene. (*A*). Primer-based construction of the N-terminus of StaC-10x. Encoded amino acids are listed below the DNA sequence, with altered amino acids and their corresponding codons highlighted in yellow. DNA sequences in bold represent restriction digestion sites, and DNA sequences in italics represent overhangs. (*B*). PCR gives rise to a fragment encoding ten modified amino acids. Top: partial scheme of PCR-based assembly, with green arrows indicating altered codons (also in green). Bottom: Result of total PCR-based assembly is a fragment with 10 altered codons (shown as green lines). (*C*). A three-part ligation incorporates the fragment into the pET28a-StaC vector, replacing the region encoding the N-terminus of the StaC protein. Restriction digest sites are indicated.

A.

```
1 --->                                                    3 --->
ggagagCATATGACCCACTCTGGCGAACGTACGGATGTT                  ggtggtccagtcggtatg
              CGCTTGCATGCCTACAAAATTAGCAACCACCACCAGGTCAGCCAT
                                                           <---   2
       M   T   H   S   G   E   R   T   D   V   L   I   V   G   G   G   P   V   G   M
```

```
                               5 --->
gcactggcgttagacttacgctaccgc       TGATTGTTTAGTTGTCGAACAGACGGAT
   gaccgcaatctgaatgcgatggcgccgtaactaacaaatcaacag
                                             <---   4
A   L   A   L   D   L   R   Y   R   G   I   D   C   L   V   V   E   Q   T   D
```

```
                               7 --->
GGCACGGTCCGTCATCC                 attggtccgcgttctatggagctgttccgc
   GTGCCAGGCAGTAGGTGCACAACCGTGGTAACCAGGCGCAAGATA
                                             <---   6
G   T   V   R   H   P   R   V   G   T   I   G   P   R   S   M   E   L   F   R
```

```
                               9 --->
cgttggggtgcagca                   GGTTGGCCAGCGGATCACCCATTAGAT
   caaccccacgtcgtctgcgttaggcattgcggccaaccggtcgcc
                                             <---   8
R   W   G   A   A   D   A   I   R   N   A   G   W   P   A   D   H   P   L   D
```

```
                               11 --->
ATTGCGTGGGTGACCAAG                tcatgagatctatcgttatcgtcgcggtacggca
   CGCACCCACTGGTTCCAGCCGCCAGTACTCTAGATAGCAATAGCA        ccgt
                                             <---   10
I   A   W   V   T   K   V   G   G   H   E   I   Y   R   Y   R   R   G   T   A
```

```
                        13 --->
gcgaatcgtcc             CCCGGAACCAGACCAAATCTGTCCGCAACACTGG
cgcttagcaggtcgcaaacaggtgtggggccttggtctggt                TGACC
                        <--- 12
   A  N  R  P  A  F  V  H  T  P  E  P  D  Q  I  C  P  Q  H  W


                        15 --->
TTAAACCCAGT             gtgggcgtgcacccagacggcccactgttattatct
AATTTGGGTCACGACTAACTCCGCCACCCGCACGTGGGTC             tgacaataataga
                        <--- 14
   L  N  P  V  L  I  E  A  V  G  V  H  P  D  G  P  L  L  L  S


                        17 --->
acgaccgtt               CGTCCAAACCGACGACCATGTCGAGGCGACGCTGACCGATCAC
tgctggcaactgccgcagcaggtttggctgct                CGACTGGCTAGTG
                        <--- 16
   T  T  V  D  G  V  V  Q  T  D  D  H  V  E  A  T  L  T  D  H


                        19 --->
GC                      cgtgcgtgcccgctttctggttgcctgcgacggcgcaagc
CGGTGGCCATGGTGGCCGTGGCACGCACGGGC                ccgcgttcg
                        <--- 18
   A  T  G  T  T  G  T  V  R  A  R  F  L  V  A  C  D  G  A  S


                        21 --->
tctcc                   GTGGTATTGAAGCCCCAGCACGCCATCGCACGCAGGTCTTC
agaggccaagcagcacggacaccataacttcggggt                TGCGTCCAGAAG
                        <--- 20
   S  P  V  R  R  A  C  G  I  E  A  P  A  R  H  R  T  Q  V  F


                        23 --->
CGTA                    gtgccccggagttaaaagatcgcctgggtgaacgtgcggcgctg
GCATTATAGAATAAGGCACGGGGCCTCAATTTT                cacgccgcgac
                        <--- 22
   R  N  I  L  F  R  A  P  E  L  K  D  R  L  G  E  R  A  A  L


                        25 --->
                        GCAGCACGTTACGCTTTCCATTACGCGCATTAAATGGTAGC
aaagtaaagaattacaattcgtcgtgcaatgcga                AATTTACCATCG
                        <--- 24
   F  H  F  L  M  L  S  S  T  L  R  F  P  L  R  A  L  N  G  S
```

```
                27 --->
GATT            gttggtgcagatgatgatacgggtgcccgtccggatgcatta
CTAAATATGGCAGACTGCCAACCACGTCTACTA            ggcaggcctacgtaat
                 <--- 26                                    <---
 D   L   Y   R   L   T   V   G   A   D   D   D   T   G   A   R   P   D   A   L
```

```
gca
cgtgACTAGTtcctgc
       < --- 28
 A
```

B.

PCR
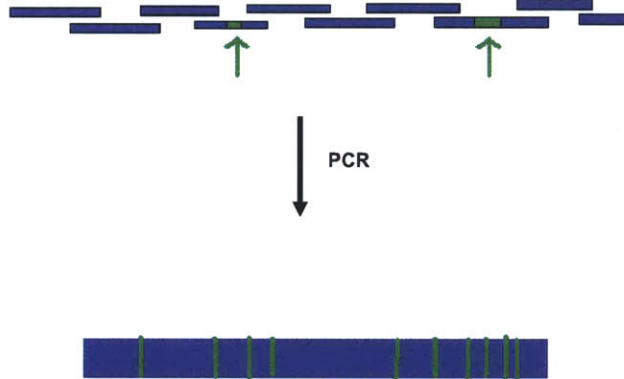
C.
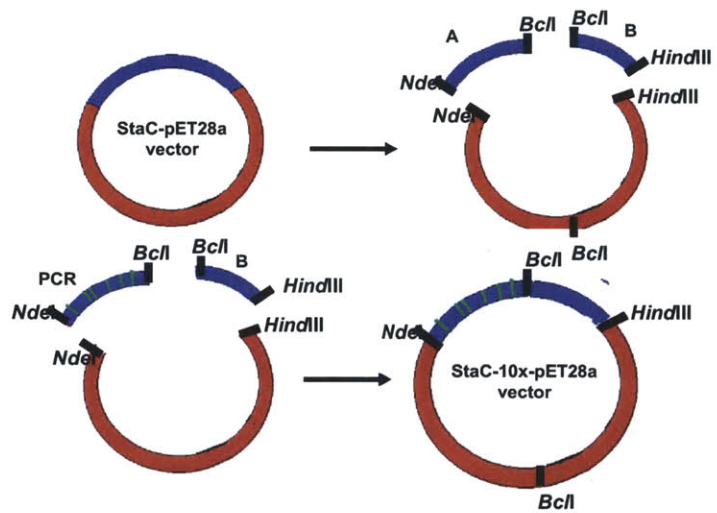
II.VII REFERENCES

1. Drennan CL & Ryan KS (2009) Divergent pathways in the biosynthesis of bisindole natural products. *Chem Biol* 16(4):351-364.

2. Jensen PR, Williams PG, Oh DC, Zeigler L, & Fenical W (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus Salinispora. *Appl Environ Microbiol* 73(4):1146-1152.

3. Sánchez C, Méndez C, & Salas JA (2006) Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. *Nat Prod Rep* 23(6):1007-1045.

4. Ruegg UT & Burgess GM (1989) Staurosporine, K-252 and UCN-01: potent but nonspecific inhibitors of protein kinases. *Trends Pharmacol Sci* 10(6):218-220.

5. Jimeno A, et al. (2008) Phase I and pharmacokinetic study of UCN-01 in combination with irinotecan in patients with solid tumors. *Cancer Chemother Pharmacol* 61(3):423-433.

6. Welch S, et al. (2007) UCN-01 in combination with topotecan in patients with advanced recurrent ovarian cancer: a study of the Princess Margaret Hospital Phase II consortium. *Gynecol Oncol* 106(2):305-310.

7. Edelman MJ, et al. (2007) Phase I and pharmacokinetic study of 7-hydroxystaurosporine and carboplatin in advanced solid tumors. *Clin Cancer Res* 13(9):2667-2674.

8. Nock CJ, et al. (2011) A phase I study of rebeccamycin analog in combination with oxaliplatin in patients with refractory solid tumors. *Invest New Drugs* 29(1):126-130.

9. Dowlati A, et al. (2009) Phase II and pharmacokinetic trial of rebeccamycin analog in advanced biliary cancers. *Cancer Chemother Pharmacol* 65(1):73-78.

10. Bailly C, et al. (1997) DNA cleavage by topoisomerase I in the presence of indolocarbazole derivatives of rebeccamycin. *Biochemistry* 36(13):3917-3929.

11. Sánchez C, et al. (2005) Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc Natl Acad Sci USA* 102(2):461-466.

12. Howard-Jones AR & Walsh CT (2006) Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. *J Am Chem Soc* 128(37):12289-12298.

13. Ryan KS, et al. (2007) Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC. *Proc Natl Acad Sci USA* 104(39):15311-15316.

14. Eppink MH, Schreuder HA, & Van Berkel WJ (1997) Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci* 6(11):2454-2458.

15. Kim SY, et al. (2007) Genetic organization of the biosynthetic gene cluster for the indolocarbazole K-252a in *Nonomuraea longicatena* JCM 11136. *Appl Microbiol Biotechnol* 75(5):1119-1126.

16. Gao Q, Zhang C, Blanchard S, & Thorson JS (2006) Deciphering indolocarbazole and enediyne aminodideoxypentose biosynthesis through comparative genomics: insights from the AT2433 biosynthetic locus. *Chem Biol* 13(7):733-743.

17. Asamizu S, Shiro Y, Igarashi Y, Nagano S, & Onaka H (2011) Characterization and functional modification of StaC and RebC, which are involved in the pyrrole oxidation of indolocarbazole biosynthesis. *Biosci Biotech Bioch* 75(11):2184-2193.

18. Groom K, Bhattacharya A, & Zechel DL (2011) Rebeccamycin and staurosporine biosynthesis: insight into the mechanisms of the flavin-dependent monooxygenases RebC and StaC. *Chembiochem* 12(3):396-400.

19. Hefti MH, Milder FJ, Boeren S, Vervoort J, & van Berkel WJ (2003) A His-tag based immobilization method for the preparation and reconstitution of apoflavoproteins. *Biochim Biophys Acta* 1619(2):139-143.

20. Massey VA (1991) Simple method for the determination of redox potentials. *Flavins and Flavoproteins*, eds Curti B, Ronchi S, & Zanetti G (Walter de Gruyter, Berlin, Germany), pp 59-66.

21. Palfey BA, Entsch B, Ballou DP, & Massey V (1994) Changes in the catalytic properties of *p*-hydroxybenzoate hydroxylase caused by the mutation Asn300Asp. *Biochemistry* 33(6):1545-1554.

22. Montersino S, Tischler D, Gassner GT, & van Berkel WJH (2011) Catalytic and structural features of flavoprotein hydroxylases and epoxidases. *Adv Synth Catal* 353(13):2301-2319.

23. Enroth C, Neujahr H, Schneider G, & Lindqvist Y (1998) The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure* 6(5):605-617.

24. Gatti DL, *et al.* (1994) The mobile flavin of 4-OH benzoate hydroxylase. *Science* 266(5182):110-114.

25. Ballou DP (2007) Crystallography gets the jump on the enzymologists. *Proc Natl Acad Sci USA* 104(40):15587-15588.

26. Ryan KS, *et al.* (2008) The FAD cofactor of RebC shifts to an IN conformation upon flavin reduction. *Biochemistry* 47(51):13506-13513.

27. Gasteiger E, *et al.* (2005) Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, ed Walker JM (Humana Press), pp 571-607.

28. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

29. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.

30. Brunger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54(Pt 5):905-921.

31. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2(11):2728-2733.

32. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213-221.

33. Badger J & Hendle J (2002) Reliable quality-control methods for protein crystal structures. *Acta Crystallogr D* 58(Pt 2):284-291.

34. Howard-Jones AR & Walsh CT (2006) Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. *J Am Chem Soc* 128(37):12289-12298.

Chapter 3.

X-ray Structure of an AdoMet Radical Activase Reveals an Anaerobic Solution for Formylglycine Posttranslational Modification

This chapter has been accepted for publication in the Proceedings of the National Academy of Sciences.

Authors:
Peter J. Goldman,[1] Tyler L. Grove,[4] Lauren A. Sites,[4] Martin I. McLaughlin,[1,4] Squire J. Booker,[4,5] and Catherine L. Drennan[1,2,3]

[1]Department of Chemistry, [2]Department of Biology, and the [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA
[4]Department of Chemistry and [5]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802 USA

III.I SUMMARY

Arylsulfatases require a maturating enzyme to perform a co- or posttranslational modification to form a catalytically-essential formylglycine (FGly) residue. In organisms that live aerobically, molecular oxygen is used enzymatically to oxidize cysteine to FGly. Under anaerobic conditions, *S*-adenosylmethionine (AdoMet, SAM) radical chemistry is employed. Here we present the first structures of an anaerobic sulfatase maturating enzyme (anSME), both with and without peptidyl-substrates, at 1.6–1.8 Å resolution. We find that anSMEs differ from their aerobic counterparts in using backbone-based hydrogen-bonding patterns to interact with their peptidyl-substrates, leading to decreased sequence specificity. These anSME structures from *Clostridium perfringens* are also the first of an AdoMet radical enzyme that performs dehydrogenase chemistry. Together with accompanying mutagenesis data, a mechanistic proposal is put forth for how AdoMet radical chemistry is co-opted to perform a dehydrogenation reaction. In the oxidation of cysteine or serine to FGly by anSME, we identify D277 and an auxiliary [4Fe-4S] cluster as the likely acceptor of the final proton and electron, respectively. D277 and both auxiliary clusters are housed in a cysteine-rich C-terminal domain, termed SPASM domain, that contains homology to ~1400 other unique AdoMet radical enzymes proposed to use [4Fe-4S] clusters to ligate peptidyl-substrates for subsequent modification. In contrast to this proposal, we find that neither auxiliary cluster in anSME binds substrate and both are fully ligated by cysteine residues. Instead, our structural data suggest that the placement of these auxiliary clusters creates a conduit for electrons to travel from the buried substrate to the protein surface.
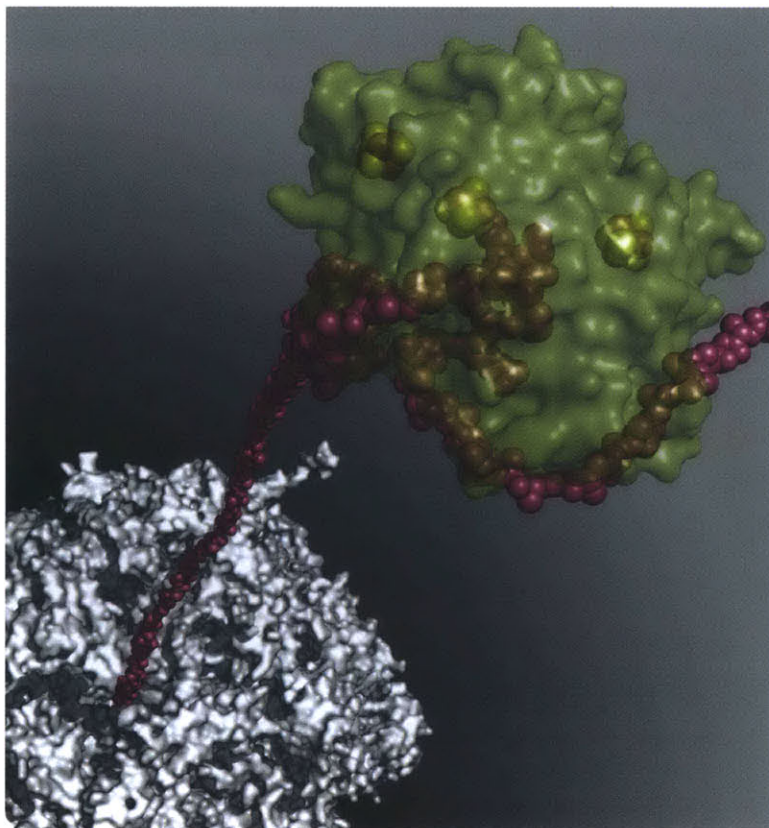
III.II INDRODUCTION

Posttranslational modification expands the chemical repertoire of enzymes, in some cases by generating modified amino acids that are well suited to perform specific reactions. Arylsulfatases, for example, require the co- or posttranslational formation of a catalytically essential formylglycine (FGly) moiety in order to perform their hydrolysis function, removing sulfate groups from a wide array of substrates (e.g. sulfated polysaccharides, sulfolipids and steroid sulfates) (1-3). In humans, lack of sulfatase activity can lead to disease (4), while in bacteria, inhibition impairs colonizing the mucosal layer of the host's gut (5). The maturation of these sulfatases involves two classes of enzymes, one that requires molecular oxygen and another that can function in its absence. Formylglycine generating enzymes (FGEs), found in eukaryotes or aerobically-living prokaryotes, generate FGly by oxidizing a cysteine residue on the target sulfatase using molecular oxygen (6, 7), whereas anaerobic sulfatase maturating enzymes (anSMEs) generate FGly from either cysteine or serine residues on their target sulfatases using S-adenosyl-L-methionine (AdoMet) radical chemistry (8-11). In addition to their importance for sulfatase chemistry, FGEs have commercial applications for generating site-specific 'aldehyde tags' to use in protein-labeling technology (12). While FGEs have been characterized in terms of structure and mechanism (6, 7), far less is known about their anaerobic cousins, the anSMEs. Here we provide the first structural insights into these unusual AdoMet radical enzymes.

The AdoMet radical enzyme family catalyzes a diverse array of radical based reactions, including sulfur insertions, complex chemical transformations and rearrangements, DNA and RNA modifications, and, in the case of anSMEs, dehydrogenation (13) (Figure III.1). Members of this family have historically been identified by a $CX_3CX\Phi C$ motif (where Φ is an aromatic residue), which ligates three of the four irons of a [4Fe-4S] cluster (14, 15), leaving the fourth iron free to bind AdoMet (16). Radical chemistry is initiated by the injection of an electron via the [4Fe-4S] cluster into AdoMet, resulting in the homolytic cleavage of the molecule into methionine and a 5'-deoxyadenosyl radical (5'dA•, Figure III.1). This radical species subsequently abstracts a hydrogen atom from substrate, resulting in 5'-deoxyadenosine (5'dA) and a substrate radical. Differentiation among the family members is a result of the action of this substrate radical. In anSMEs, the removal of a proton and an electron from the radical intermediate completes catalysis (10, 17) (Figure III.1).

The AdoMet radical dehydrogenase subfamily includes anSMEs and the carbohydrate natural product biosynthetic enzyme BtrN (18, 19). Interestingly, both enzymes harbor additional [4Fe-4S] clusters that are necessary for turnover (20). In the case of BtrN, one

auxiliary cluster has been identified (21), while anSMEs have two (10, 17). For anSME, the sequence surrounding these two clusters, including a previously identified 7-cysteine motif ($CX_{9-15}GX_4C$—gap—$CX_2CX_5CX_3C$—gap—C) (17), places it in a ~1400-membered AdoMet radical subfamily that was recently described by Haft and Basu through bioinformatic analysis and thought to function in the modification of ribosomally translated peptides (22). This subfamily has been designated TIGR04085 and named SPASM for its biochemically characterized founding members AlbA, PqqE, anSMEs, and MtfC, which are involved in subtilosin A, pyrroloquinoline quinone, anaerobic sulfatase, and mycofactocin maturation respectively (22, 23). While the function of these auxiliary clusters is unknown, the 7-cysteine motif prompted speculation that members of the SPASM subfamily, including anSMEs, use an available ligation site on one of the [4Fe-4S] clusters for substrate binding (10). Direct binding of substrate to an auxiliary [4Fe-4S] cluster has been observed in the molybdenum cofactor biosynthetic enzyme MoaA, another AdoMet radical enzyme (24, 25). Other possible functions specific to anSMEs include the second oxidation of substrate or substrate deprotonation (Figure III.1).

In this work, we report structures of anSMEcpe, a biochemically characterized anaerobic sulfatase maturating enzyme from *Clostridium perfringens*, which can oxidize either serines or cysteines into FGly sidechains (26, 27). We have solved the structure of a His_6 tagged protein construct with AdoMet bound and three structures of an untagged protein construct in the presence of AdoMet; one substrate free and two others with different peptide substrates bound. The overall structure and substrate binding characteristics of anSMEcpe differ significantly from the aerobic FGE system, which clarifies the difference in promiscuity between the two enzyme families. Surprisingly, the structures show full cysteine ligation of both auxiliary clusters, which has important implications for the anSMEcpe mechanism, as well as for SPASM family members as a whole.

III.III RESULTS

Iron anomalous signal from a dataset collected at a home $Cu-K_\alpha$ source was used to solve an initial structure of His_6 tagged anSMEcpe ('His_6, AdoMet-bound' structure; Table III.1). Due to occupancy of the His_6 tag in the substrate binding site, additional structures of native, untagged anSMEcpe were solved with and without two substrate peptides ('AdoMet-bound', 'Kp18Cys, AdoMet', and 'Cp18Cys, AdoMet' structures; Table III.1). All structures of anSMEcpe contain three [4Fe-4S] clusters including the AdoMet cluster housed in the N-terminal AdoMet radical domain and two auxiliary clusters located in the C-terminal SPASM domain. The two

domains are connected by the α6a helix and the protein terminates with α6' helix (Figure III.2.A-C).

Structural features of the anSMEcpe AdoMet radical domain

The N-terminal domain of anSMEcpe is a parallel $(\beta/\alpha)_6$ partial TIM barrel, spanning residues 3–234 (Figure III.2.A and III.2.C, magenta). This fold is common to nearly all other structurally characterized members of the AdoMet radical superfamily (28, 29). We will refer to this partial barrel as the AdoMet domain. Here, the AdoMet radical sequence motif $(CX_3CX\Phi C)$ is found in a loop following the β1 strand. C15, C19, and C22 each ligate an iron atom in one of the [4Fe-4S] clusters, referred to here as the AdoMet cluster. The fourth, so-called 'unique iron,' is ligated by the amine nitrogen and carboxyl oxygen from the methionine moiety of AdoMet, as expected (16). All structures contain clear density for AdoMet bound in the active site (Figure III.3.A), with the exception of a substrate bound (Cp18Cys) cocrystal structure where the shape and position of density for AdoMet in chain B is minimal and inconsistent with other models.

Four previously described AdoMet binding motifs are conserved in anSMEcpe, including the 'GGE' motif, the ribose motif, the 'GXIXGXXE' motif, and the β6 motif (Figures III.3.C and III.4) (28, 29). In addition, the backbone of Y21 (the hydrophobic residue in the AdoMet radical $CX_3CX\Phi C$ motif) hydrogen bonds with the N6 position of adenine and R143, just following α4a, stabilizes the ribosyl and carboxyl moieties of AdoMet. An arginine following α4a makes a similar interaction in the AdoMet radical proteins HemN and HydE (30, 31). On the backside of the AdoMet domain, a patch of conservation can be found following the β2 strand (Figure III.3.B). This site is the proposed binding location of the physiological reductant, commonly flavodoxin (30, 32).

The SPASM domain and auxiliary cluster binding

The α6a helix links the AdoMet domain to the C-terminal SPASM domain (Figure III.2.A-C, green), the latter containing both auxiliary clusters. According to the designation by TIGR04085, the SPASM domain initiates at C261, not C255 (the first Fe ligating cysteine in anSMEcpe). However, both cysteines are conserved in anSMEs and it is common for SPASM domain-containing proteins to have a proximal upstream cysteine (22). In the anSMEcpe structures, C255 and C261 ligate the first auxiliary cluster, Auxiliary Cluster I (Aux I), before the

backbone folds into a beta hairpin (V266–Y274). Residues of this hairpin exhibit high conservation to other members of the SPASM family (Figure III.2.D). Immediately following the hairpin, C276 provides the third ligand to Aux I. A variable alpha helical region follows, providing a barrier between this cluster and solvent. Aux I lies 16.9 Å from the AdoMet cluster (measured from the closest atom in each cluster).

The second portion of the SPASM domain contains the $CX_2CX_5CX_3C$ part of the 7-cysteine motif. The first three cysteines (C317, C320, and C326) provide three ligands to the second auxiliary cluster, Auxiliary Cluster II (Aux II), while the protein backbone forms small helical interactions that surround the cluster. The fourth cysteine of the motif, C330, crosses back to provide the final ligation site to Aux I (Figure III.2.D). Very high sequence conservation is found in the linear CXXXC region bridging the two clusters, where the last two residues before the final cysteine of the motif are glycines (G328 and G329 in anSMEcpe, Figure III.2.D). Aux II lies 12.9 Å away from Aux I and 26.7 Å away from the AdoMet cluster (Figure III.2.B). A series of loops follows $CX_2CX_5CX_3C$ before C348 (the final of the seven cysteines) occupies the final ligation site of Aux II, ending the SPASM domain and initiating the α6′ helix. This helix lies adjacent to the α6 helix and completes the barrel. The C-terminus of the protein lies at the end of this helix (Figure III.2.C).

Structural homology to MoaA

While the AdoMet domain of anSMEcpe is very similar to other members of the AdoMet radical family, the C-terminal SPASM domain is structurally similar (RMSD 6.3 Å) to only one other AdoMet radical protein, MoaA (33). Like anSMEcpe, MoaA ligates a C-terminal auxiliary cluster that overlays well with anSMEcpe's Aux I (Figure III.5). Following two cysteine ligands to its auxiliary cluster, MoaA has a beta hairpin that shares high sequence homology with anSMEcpe and other SPASM domain containing proteins, including G273 (G271 in anSMEcpe) in the n+3 position of the hairpin turn and Y276 (Y274 in anSMEcpe), a residue that contributes to a hydrophobic pocket adjacent to the hairpin. MoaA then has a third cysteine ligand and terminates after a helical region. It lacks both the $CX_2CX_5CX_3C$ motif and the second auxiliary cluster that are common in SPASM family members. With only these three protein ligands to its auxiliary cluster, MoaA uses the available coordination site to bind substrate (24, 25). In anSMEcpe, the final cysteine of the $CX_2CX_5CX_3C$ motif is the fourth ligand to this cluster (Figure III.5.E-F).

A cavity underneath the barrel is the only access to the active site from the exterior of the protein. In all structures of anSMEcpe solved using the His_6-anSMEcpe construct, multiple histidines of the C-terminal tag could be modeled into residual electron density. These residues appear to block access to the active site. However, when native protein was used for cocrystallization with AdoMet and peptide substrates, peptide density reaching into the active site was present. Two substrate bound structures were solved using peptides designed to mimic a *C. perfringens* (Cp18Cys) and a *Klebsiella pneumoniae* (Kp18Cys) sulfatase protein (see Figure III.6.A for sequences). Each structure contains density for 9-11 residues of the 18mers (Figure III.7.A). These peptide residues stretch from the exterior of the protein to the active site, both entering and exiting via the cavity at the bottom of the barrel (Figure III.8.A-B). Buried surface area of the protein – peptide interaction for the four bound peptides to anSMEcpe (two per asymmetric unit) is 786 ± 38 $Å^2$, corresponding to 60.7 ± 1.7 % of the total surface area of the modeled portion of the peptides (34). This binding mode differs from the aerobic sulfatase maturating enzymes, FGEs, which use the same CXPXR motif for substrate recognition. In the aerobic system, the sulfatase maturase adopts a very different fold that lacks the internal cavity found in anSMEcpe. Instead, peptide binding and catalysis occurs on a surface exposed region of the enzyme (Figure III.8.C) (7). Further, in the aerobic system, only 4 of the 12 hydrogen bonds between FGE and its substrate use the substrate backbone (PDB ID 2AIJ). The majority of interactions are made between peptide sidechains and the maturase, explaining the high sequence specificity in this system (7).

Other than the arginine residue of the CX(A/P)XR motif (Figure III.9.A-B), this sidechain-based sequence specificity is not seen in anSMEs. First, anSMEcpe is able to accommodate considerable substrate sequence variation on either side of the target cysteine (Figure III.9.C-D), allowing the two peptides (Cp18Cys and Kp18Cys) to bind in an almost identical orientation (Figure III.5.A). In both cases, backbone hydrogen bonds to anSMEcpe are the primary means of stabilization (Figure III.7.C). Only two positions, 4 and 10, of the peptides differ, resulting in two additional hydrogen bonds at the 4 position of Kp18Cys (Figures III.6-7). The remaining 17 hydrogen bonding interactions are conserved between the two substrates. Of these, 12 are formed between peptide backbone and anSMEcpe and 5 are formed between anSMEcpe and a single peptidyl sidechain, R11 of the conserved CX(A/P)XR sulfatase motif. The extensive binding pocket created for this arginine, made up of F188, E159, and L118, uses π stacking, electrostatics/hydrogen bonding, and van der Waals interactions, respectively (Figure

III.9.A). These interactions appear to be the anchor for the peptide and are conserved among the other biochemically characterized anSMEs (Figure III.4).

Identification of catalytic residues

Aided by two prolines, the substrate peptide makes a tight turn in the active site, allowing the target cysteine to protrude into the deepest part of the barrel, just below AdoMet. The cysteinyl $C\beta$ is located 4.1 Å from the 5′ carbon of AdoMet (Figure III.10.A). This distance is in agreement with previously reported distances between the 5′ position of AdoMet and the substrate hydrogen abstraction site (3.8–4.1 Å) (28). The orientation of the cysteine directs the $C\beta$ *pro-S* hydrogen toward the AdoMet 5′ position, matching biochemical evidence for the enzyme's stereoselectivity (26).

During catalysis, a general base is needed for deprotonation of the cysteine sidechain to allow the formation of the thioaldehyde (Figure III.1). Analysis of the peptide-binding pocket revealed the presence of two residues with titratable sidechains within 5 Å of the substrate cysteinyl sulfur position, D277 and Y24 (Figure III.10.A). Y24 is two residues downstream of the AdoMet radical domain's $CX_3CX\Phi C$ motif, while D277 is in the SPASM domain and adjacent to C276, an Auxiliary Cluster I ligand. To identify the catalytic residue, two mutants were generated and assayed for activity, anSMEcpe D277N and Y24F. Compared to wild type, the Y24F mutant retains 11.7% FGly production activity, while the D277N mutant only retains 0.8% activity. Along with this decreased activity, an uncoupling of the production of FGly and 5′dA is observed for the D277N mutant (Figure III.10.B). The proximity to the substrate sulfur and a large decrease in activity imply a catalytic role for the D277 sidechain.

Sidechain movements upon substrate binding

Proximal to the catalytic residue D277 are the positions of two glutamine residues that are the only two residues to undergo rearrangement when substrates bind (Figure III.11). When AdoMet is bound, but no peptide is present (AdoMet structure), the sidechain of Q98 points into the middle of the barrel (Position 1 in Figure III.11). In this orientation, Q98 would clash with multiple sites on the peptide backbone. In chain B of the Cp18Cys peptide structure (where AdoMet is disordered), Q98 points towards the AdoMet cluster (position 2) where it would form unfavorable close contacts between its amide and the $C\alpha$, $C\beta$, and $C\gamma$ carbons of an

ordered AdoMet methioninyl moiety. In both chains of the Kp18Cys structure, Q98 is bent behind AdoMet, hydrogen bonding with the cofactor's carboxyl group (Position 3 in Figure III.11). In this position of Q98, the binding of both AdoMet and peptide are favored.

Similar to Q98, Q64 has distinct orientations depending on the composition of the active site. First, when no peptide is bound, this sidechain also points toward the center of the barrel, clashing with the sidechain position of peptide residue 8 in an overlaid structure (darker color in Figure III.11). The geometry of the sidechain in this position is routinely flagged as 'generously allowed' in Ramachandran analysis. In chain B of the Cp18Cys structure (disordered AdoMet), Q64 interacts with both K9 and D277, the putative catalytic residue. However, in this structure only one hydrogen bond is formed with D277, and as a result, D277 is found with partial occupancy in two orientations (lighter colors in Figure III.11). Also in this structure, the cysteine sidechain of substrate is pointed away from AdoMet, in a noncatalytic orientation where the *pro-R* hydrogen of the cysteine Cβ is proximal to the AdoMet binding site. In the Kp18Cys structure, where all substrates are present, Q64 makes two hydrogen bonds with D277. Here, only one orientation of D277 (the 'catalytic' orientation) is observed and the substrate cysteine is also in the catalytic orientation, with the *pro-S* hydrogen of the cysteinyl Cβ closest to the 5' position of AdoMet (Figure III.11.A-B). Thus the binding of AdoMet and peptide appear to re-orient both Q98 and Q64, which in turn, stabilize AdoMet, D277, and the target Cys in catalytic conformations.

Electron transfer pathway

After hydrogen abstraction and thiol deprotonation, formation of the thioaldehyde drives oxidation of the Cβ, requiring an electron acceptor to complete catalysis. Cβ lies 8.6 and 8.9 Å from Aux I and the AdoMet cluster, respectively, indicating both clusters are within suitable distance to be electron transfer partners (35). If an electron transfer event results in the reduction of Aux I, removing the electron from the system would likely require transfer from Aux I to Aux II, as peptide binding provides a barrier between Aux I and solvent (Figure III.12). Aux I lies near the bottom opening of the anSMEcpe barrel. When no peptide is present, the cluster is 9.7 Å from bulk solvent, with the inside of the protein barrel as the closest protein – bulk solvent interface (Figure III.12.A). In this conformation, the cluster has a similar residue depth to both the AdoMet cluster (9.5 Å) and Aux II (8.3 Å). However, when substrate is present this avenue to solvent is cut off, and the shortest path to bulk solvent is below the barrel, 11.0 Å

away (Figure III.12.B). Peptide binding does not affect the residue depth of either the AdoMet cluster or Aux II.

III.IV DISCUSSION

Here we present the first structures of an anaerobic sulfatase maturating enzyme, anSMEcpe, which allows us to compare how nature evolved anaerobic as well as aerobic solutions for the same enzyme function. While both enzyme classes are designed to bind their target sulfatase, in one case, binding must involve sequestering a Cys/Ser to afford radical based chemistry, while in the other, the target Cys must be accessible to interact with molecular oxygen (6-8, 17). We find that the aerobic FGEs and anSMEs use different protein folds, with the N-terminal domain of anSMEs sharing a classic AdoMet radical partial barrel fold (28, 29), consistent with the chemistry being performed. While the active site of FGE is on the surface of that enzyme, where it is readily accessible to molecular oxygen, the active site of anSME is buried in a cleft created between the C-terminal SPASM domain and the N-terminal AdoMet radical domain. To fit into this cavity, the target peptide adopts a relatively tight turn, perhaps explaining the preference for Ala or Pro in position of the conserved (S/C)X(A/P)XR motif. In contrast, peptides bound to FGE have no apparent conformational restraints (Figure III.8) (7).

FGEs and anSMEs also vary in their substrate selectivity. Compared to FGEs, anSMEs are able to act on a larger variety of peptide substrates. For example, anSMEcpe itself can bind and catalyze FGly formation on a *C. perfringens* substrate analog as well as a *K. pneumoniae* substrate analog (26). Another anSME from *Bacteroides thetaiotaomicron* is responsible for activating up to 28 sulfatases under anaerobic conditions (5). From structural comparisons, we can now explain this substrate specificity variation between these enzyme classes. While the aerobic system uses a mainly sidechain – maturase hydrogen bonding network for substrate stabilization, the anSME system uses a primarily backbone – maturase hydrogen bonding network, with only the Arg of the (C/S)X(A/P)XR motif involved in sidechain-based hydrogen bonding. This reliance by SMEs on primarily peptide backbone-based hydrogen bonding interactions is in agreement with the much higher degree of promiscuity that exists in the anSME system in relation to the FGE system. Interestingly, pyruvate formate lyase activating enzyme (PFL-AE), the only other structurally characterized AdoMet radical enzyme involved in protein modification, also uses primarily peptide backbone – activase interactions (32). As more structures become available, it will be interesting to see if this binding mode will be common to AdoMet radical enzymes that act on protein substrates.

These structures of an anSME also provide insight into the catalytic mechanism of this enzyme class. For catalysis, at least five steps are required (Figure III.1). For the first step of 5'dA• generation, the high degree of similarity between the structure of the anSME AdoMet radical domain and structures of other AdoMet radical enzymes suggests that all components necessary for radical generation are found in this N-terminal domain, implying that anSMEs share a common initiation mechanism with the rest of the AdoMet radical superfamily.

In terms of substrate radical generation, these structures show the Cβ of the target Cys 4.1 Å from the AdoMet 5'C, in agreement with all other structures of AdoMet radical enzymes that have substrates bound (Figure III.10). Very little movement within anSMEcpe is required for substrate to bind in this catalytic position. Only two glutamines, which are well conserved in anSMEs (Figure III.4 blue), have a distinct orientation in the presence or absence of either AdoMet or peptide substrate. These residues aid both the stabilization of AdoMet and in the positioning of D277 (Figure III.11).

In contrast to radical generation, substrate deprotonation is only required by a small number of AdoMet radical enzymes. Before this work, it was not clear in anSME whether auxiliary cluster(s), enzyme residue(s), or both are involved in this reaction step. Here, the structures of anSME with peptides bound show that the substrate does not directly ligate either auxiliary cluster as previously suggested (10), and that both auxiliary clusters are far from the target Cys (8.6 Å and 20.8 Å), making it highly unlikely these clusters are involved in this deprotonation and/or inner-sphere electron transfer. Instead, the structure reveals two residues with titratable sidechains that are close to the target Cys (Y24 from the AdoMet radical domain and D277 from the SPASM domain), and mutagenesis studies are consistent with D277 as the catalytic base (Figure III.13). It is interesting that the SPASM domain, and not the AdoMet radical domain, contributes this key catalytic residue that differentiates anSME's chemistry from that of other AdoMet radical enzymes, as this finding is consistent with previous structural studies that also showed the importance of residues outside of the partial barrel radical fold to the diversification of AdoMet radical chemistry (28, 29).

The next step, substrate oxidation, is again only required by certain subfamilies of AdoMet radical enzymes such as the dehydrogenases studied here, and others, like the heme biosynthetic enzyme HemN (30). While the Aux I and AdoMet clusters are nearly equidistant and within acceptable electron transfer distances (8.6 and 8.9 Å) from the substrate Cβ (Figure III.8), we propose that Aux I is the immediate electron acceptor for this oxidation (Figure III.13). With few exceptions (36, 37), reduction of the AdoMet cluster during catalysis has only been

proposed in systems that, unlike anSMEcpe, use AdoMet catalytically. Assuming that Aux I is the electron acceptor, for a subsequent turnover, it would need to be reoxidized. Here we further propose that Aux II, 12.9 Å away, performs this function (Figure III.13). Other options for the reoxidation of Aux I are more problematic: the AdoMet cluster is too far from Aux I for direct electron transfer (16.9 Å) and the closest protein surface to which an external electron acceptor could bind appears blocked by bound substrate (Figure III.12). Thus, Aux II is the most viable candidate.

We can further consider if electrons are re-cycled in this reaction, i.e. an electron used to homolytically cleave AdoMet in one cycle is derived from a previous cycle's substrate oxidation. By monitoring the level of flavodoxin semiquinone depletion during anSMEcpe catalysis, Grove et al. have recently established that an electron can indeed be re-cycled in this fashion (26). While Aux II is also too far from the AdoMet cluster for direct electron transfer, an external electron acceptor, like flavodoxin, could accept an electron from Aux II and redeposit it into the AdoMet cluster. Electrochemical characterization of all three clusters would provide validation that i) Aux I is the substrate radical electron acceptor during catalysis and ii) that electron transfer between Aux I and Aux II is possible. In the meantime, the structures described here reveal distances that support a role for these clusters in substrate oxidation by electron transfer, and refute other possible functions, including a role in substrate binding and deprotonation, as discussed above.

In addition to the mechanistic insight provided by these structures, visualization of the C-terminal domain of anSMEcpe clarifies the function of the recently described SPASM domain. Accession TIGR04085 designates 281 sequences as SPASM subfamily members; however, when these sequences are input in the Enzyme Function Initiative's recent AdoMet radical superfamily clustering effort (http://enzymefunction.org/resources/workshops), these 281 sequences are found in 153 different nodes containing 1,392 unique sequences (Figure III.14). While this is the first structure of a SPASM domain-containing enzyme, we find that the first part of the domain, containing two of the seven cysteines of the 7-Cys motif, has been visualized before in the structure of the AdoMet radical protein MoaA. In particular, anSMEcpe shares with MoaA a conserved beta hairpin that extends the AdoMet radical beta sheet and contains cysteines on either end of the turn (Figure III.2). In both cases, these two cysteines ligate an auxiliary [4Fe-4S] cluster along with a third upstream cysteine (C255 in anSMEcpe) (Figure III.5). This beta hairpin also has high sequence homology to non-SPASM AdoMet radical dehydrogenase BtrN. Like MoaA, BtrN contains an auxiliary [4Fe-4S] cluster and, while BtrN has not been structurally characterized, the MoaA auxiliary cluster superimposes very well with

Aux I in anSMEcpe (Figure III.5). Thus, this beta hairpin motif that is flanked by cysteines appears to be associated with binding an auxiliary [4Fe-4S] cluster in more than just SPASM-domain containing proteins. Since this hairpin is a subdomain of SPASM, we will refer to it as a twitch subdomain.

Following the twitch subdomain, MoaA completes two helices and terminates, while anSMEcpe continues the SPASM domain with one helix leading to the next set of four cysteines arranged in a $CX_2CX_5CX_3C$ motif. Interestingly, while all four cysteines of this motif bind to an auxiliary cluster, they do not ligate the same cluster. The first three cysteines ligate Aux II, while the final cysteine ligates Aux I. Therefore Aux I is coordinated by one upstream cysteine, two twitch subdomain cysteines, and one cysteine from the SPASM domain's $CX_2CX_5CX_3C$ motif. MoaA shares this cluster coordination except for the last Cys; this Fe site is available for substrate binding (Figure III.5.C,F). In anSME, the bridging sequence between the final two cysteines of $CX_2CX_5CX_3C$ is Lys-Gly-Gly. This 'XGG' sequence is highly conserved in TIGR04085 and could function to ensure a viable electron transfer pathway environment between the two auxiliary clusters. The final Cys of the 7-Cys motif coordinates Aux II.

Cysteine ligation of the auxiliary clusters was accurately predicted by TIGR04085, which establishes seven cysteine ligands for the two auxiliary clusters. However, an upstream cysteine, C255 in anSMEcpe, is also involved in cluster binding, resulting in the unexpected full protein ligation of both clusters. This result refutes the idea that all SPASM domains have an available ligation site for substrate binding and indicates that at least anSMEs do not use an auxiliary cluster for this purpose. An analysis of TIGR04085 reveals that 37% of SPASM domain proteins have a cysteine that is both upstream of the SPASM domain and downstream of the AdoMet radical domain, indicating that full ligation of the auxiliary cluster may be common among members of the SPASM subfamily. Without the role of substrate ligation, these SPASM family members may use these clusters to facilitate electron flow in or out of the active site during turnover, insinuating that their mechanisms involve some kind of redox chemistry.

In the case of SPASM family members lacking an upstream cysteine, the anSMEcpe structure indicates how substrates might coordinate Aux I. In the absence of a cysteine equivalent to C255, an available iron coordination site would be exposed to this substrate binding region (Figure III.5.C). Importantly, due to the distance between the ligation site and the AdoMet binding site, the auxiliary cluster binding and hydrogen abstraction locations must be distal (in anSMEcpe, C255 is 12 Å from the substrate hydrogen abstraction site). In the case of MoaA, where substrate does directly ligate cluster, the ligation site (the N1 position of the guanine base) is on the opposite end of the GTP substrate than the hydrogen abstraction site

(the 3' hydrogen atom of the ribose; Figure III.5.C, grey) (24, 25, 38). A similar mode of binding would be required for any SPASM members that use cluster ligation for substrate binding. As substrates in the SPASM family are predicted to be peptides, this would entail cluster ligation ~3 residues up or downstream of the hydrogen abstraction site.

In summary, the anSMEcpe system provides a great example of the modularity of the AdoMet radical superfamily. The anSMEcpe AdoMet domain has very similar structural folds and cofactor binding motifs to the rest of the superfamily. However, the end reaction catalyzed is unique from all other structurally characterized AdoMet radical proteins. While the AdoMet domain must provide all residues necessary for radical generation, it is the addition of the SPASM domain, which includes both auxiliary clusters and the catalytic residue, that steers catalysis following 5'dA• generation. Interestingly, aside from the anSMEs, the only other biochemically characterized AdoMet radical dehydrogenase, BtrN, only contains one additional [4Fe-4S] cluster. Upon structural characterization, it will be interesting to compare this enzyme to anSMEcpe and the full SPASM domain architecture.

III.V   MATERIALS AND METHODS

Cloning of the cpe0635 gene from Clostridium perfringens

The gene corresponding to anSMEcpe (*cpe0635*) from *C. perfringens* was cloned into pET26b as previously described (26) to afford a protein with a C-terminal hexahistidine tag. To generate a plasmid that produces the native form of anSMEcpe, the above gene was amplified with PCR in combination with a forward primer containing an *Nde*I restriction site (underlined) (5'-CGC-GCC-CGC-ATA-TGC-CAC-CAT-TAA-GTT-TGC-TTA-TTA-AGC-3') and a reverse primer containing a *Xho*I restriction site (underlined) (5'-CCA-CTC-GAG-TTA-TTT-AAT-ATT-GTT-GGC-AAC-ATT-TAT-TAA-CC-3'). The reverse primer also contains the natural stop codon of the *cpe0635* gene (bold). The PCR product was digested with the appropriate enzymes in parallel with pET26b. The native *cpe0635* gene was ligated into pET26b and confirmed by DNA sequencing.

Construction of D277N and Y24F variants of anSMEcpe

anSMEcpe variants   D277N   and   Y24F   were   constructed   using   the   Stratagene QuikChange II site-directed mutagenesis kit with the following primers: D277N forward primer

(5'-GGG-AGT-GTT-TAT-CCT-TGT-AAT-TTT-TAT-GTT-TTA-GAT-AAA-TGG -3'); D277N reverse primer (5'-CCA-TTT-ATC-TAA-AAC-ATA-AAA-ATT-ACA-AGG-ATA-AAC-ACT-CCC-3'); Y24F forward primer (5'-GCA-CTT-ATT-GTT-TTT-TTC-ATT-CTT-TAA-GTG-3'); Y24F reverse primer (5'-CAC-TTA-AAG-AAT-GAA-AAA-AAC-AAT-AAG-TGC-3'). All mutations were confirmed by DNA sequencing at the Pennsylvania State University Nucleic Acid Facility. Expression and purification of the variant constructs were done as described previously (26).

Expression and Purification of Native anSMEcpe

Over production of the anSMEcpe native construct was carried out as previously described for the C-terminal hexahistidine tag anSMEcpe protein (26). Purification of native anSMEcpe typically began with ~ 40 g of cell paste that was resuspended in 200 mL of lysis buffer (50mM HEPES pH 7.5, 300mM KCl, 5% glycerol, 10mM DTT) containing 1 mM PMSF, lysozyme at a final concentration of 1 mg mL^{-1} and DNase I at a final concentration of 1 μg mL^{-1}. After being stirred at room temperature for ~ 30 min, the suspension was placed in an ice bath and cooled to ~ 0 °C. After 10 min, the lysis solution was subjected to six-45 s bursts of sonic disruption at 30% output with intermittent pausing for 8 min to maintain a temp less than 8 °C. The lysate was then sealed in centrifugation bottles before being removed from the glove box and centrifuged for 30 min at 50,000 x g at 4 °C. After centrifugation, the bottles were brought back into the glove box. The supernatant (~210 mL) from the lysis was placed in a 250 mL bottle on ice with medium stirring to maintain the temperature at ~ 0 °C. A 40% ammonium sulfate precipitation was performed by adding 48.6 g of $(NH_4)_2SO_4$ slowly to the stirring solution over 30 min. After addition, the suspension was stirred for an additional 50 min before being placed in the original centrifugation bottles, sealed, and spun at 50,000 x g for 30 min. The supernatant (~ 235 mL) was then poured into another 250 mL bottle, and again cooled to ~ 0 °C on an ice bath. An additional 30 g of $(NH_4)_2SO_4$ was added to bring the concentration to 60%. After addition, the suspension was again stirred on ice for 30 min before being placed in clean centrifugation tubes, sealed, and spun at 5,000 x g for 30 min. The supernatant from the 60% cut was discarded. The pellets were stored overnight in an LN$_2$ Dewar. After storage, the pellets were resuspended in 20 mL of 25 mM Tris-HCl, pH 7.5. The protein was centrifuged, and then concentrated to ~ 10 mL with an Amicon stirred cell equipped with a 10 kDa MWCO membrane.

The protein was subsequently fractioned over a 16/26 Sephacryl S-200 column equilibrated in 50mM Tris-HCl pH 7.5, 500 mM NaCl, 5 % Glycerol, 5 mM DTT. Fractions from

approximately 55 mL to 75 mL were combined, concentrated with an Amicon Centricon with a 10 kDa MWCO membrane, and finally diluted with 25 mM Tris-HCl to bring the NaCl concentration to ~ 100 mM. The protein was then passed over a HiPrep Q FF 16/10 column equilibrated in 50mM Tris-HCl pH 7.5, 100mM NaCl, 5 mM DTT. The anSMEcpe that eluted in the void volume was combined and reconstituted as previously described (26). The protein was then concentrated, buffer exchanged with a PD-10 column equilibrated in 50mM HEPES pH 7.5, 500 mM KCl, 10% Glycerol, 5 mM DTT, and fractioned again over a 16/26 Sephercryl S-200 column as above. Fractions corresponding to anSMEcpe were pooled and concentrated. Final yield of protein was approximately ~ 1 mg/ g of cell paste.

Crystallization

To obtain crystals of the His_6 anSMEcpe construct (Table S1), a solution containing 10 mg/mL His_6 tagged protein, 10 mM HEPES pH 7.5, 1 mM AdoMet (New England BioLabs), and 0.5 mM Cp18Cys was incubated overnight at room temperature in an anaerobic environment (95% Ar, 5% H_2, COY Laboratory Products, Inc.). Using the hanging-drop vapor diffusion method at room temperature, 2 µL of this protein solution were mixed with 2 µL precipitant (6–11% PEG 2000 and 100 mM MES, pH 6.0) and 0.4 µL of 7.0 mM LysoFos Choline 12 (Hampton Research) over a reservoir of 0.5 M LiCl. Rod shaped crystals would appear after three days and grow to dimensions of 80 x 80 x 200-400 µm. Crystals were cryo-cooled by direct submersion in liquid nitrogen.

Crystals of the native protein were grown from a solution containing 10 mg/mL untagged protein, 10 mM HEPES pH 7.5, 1 mM AdoMet (New England BioLabs), and either no peptide, 0.5 mM Cp18Cys, or 0.5 mM Kp18Cys, then incubated overnight at room temperature in an anaerobic environment (Table S1). Crystal drops were set up with 1 µL of this protein solution and 1 µL precipitant (24–34% PEG 4000, 0–300 mM ammonium acetate, and 100 mM sodium acetate, pH 4.5). These crystals would grow out of precipitate in 1–2 weeks. Crystals were harvested and washed with a cryoprotectant solution containing all appropriate salt concentrations supplemented with 15% glycerol, then submerged in liquid nitrogen. With both His_6 tagged and native protein constructs, crystals were only obtained in the presence of AdoMet.

Using an in-house Cu-K$_\alpha$ rotating anode source (Rigaku) with an imaging plate (RAXIS IV, Rigaku), a dataset of a His$_6$ crystal was collected at 100 K using sequential 1° oscillations. At this wavelength, phenix.xtriage estimated iron anomalous signal to extend to 2.6–3.0 Å. The AutoSol Wizard (39) was used to phase the dataset. Twenty-four of twenty-four iron sites (corresponding to the three [4Fe-4S] clusters of the two monomers in the asymmetric unit) were found with a figure of merit 0.422 (to 2.57 Å resolution). The experimental maps were density modified (solvent flattened and 2-fold non-crystallographic symmetry averaged) by RESOLVE and used to build an initial model of anSMEcpe using the structure of MoaA (PDB ID 1TV8) as a guide (33).

After 369 residues and most sidechains were built using the density modified experimental maps, the high resolution His$_6$ structure of anSMEcpe was solved by isomorphous replacement. Subsequent structures of native anSMEcpe complexes (AdoMet-bound, Kp18Cys-AdoMet, and Cp18Cys-AdoMet) were also solved by isomorphous replacement using the His$_6$ structure as an initial model. These datasets were collected on beamlines 24-ID-E and 24-ID-C of the Advanced Photon Source and on X-29 of the National Synchrotron Light Source. All diffraction data were integrated and scaled using the HKL2000 suite (40). COOT (41) and Phenix (42) were used for iterative rounds of model building and structure refinement, respectively, using no sigma cutoff or NCS restraints. Composite omit maps were used to verify the final model. See Table III.1 for full data processing, refinement, and validation statistics.

In all crystal forms, anSMEcpe crystallized as a monomer with very little interaction between the two molecules in the asymmetric unit or crystal lattice. The His$_6$, AdoMet bound structure is missing residues 1, 31-32, and 370 in chain A and 1, 28-32, and 370 in chain B of anSMEcpe's 370 residues due to disorder. This structure also contains residues 8-22 (VDKLAAALEHHHHHH) of the C-terminal His$_6$ tag in chain A and one well ordered histidine in chain B. The AdoMet-bound (untagged, no peptide) structure is missing residues 1-2 and 29-32 in chain A and residues 1-2 and 370 in chain B. The Kp18Cys, AdoMet structure is missing residues 1 in chain A and B and 27-32 and 370 in chain B. Also missing are residues 1-2 and 14-18 of the Kp18Cys 18mer in chain A and residues 1-2 and 13-18 of the peptide in chain B. The Cp18Cys, AdoMet structure is missing residues 1 and 26-33 in chain A and B and residue 34 and 370 in chain B. Also missing are residues 1-2 and 13-18 in both chains the Cp18Cys 18mer. Chain B of this anSMEcpe structure is the only monomer in which AdoMet is not modeled into the visible in the electron density; however, the peptide density in this chain is superior. In addition, only two monomers (chain B of the AdoMet-bound and chain A of the Kp18Cys,

AdoMet structures) contain residues 29-32, a disordered loop immediately following the AdoMet binding cluster motif. All models are fully constituted with three [4Fe-4S] clusters.

Activity determination of anSMEcpe

Activity assays with Kp18Cys peptide were carried out as described (26). Activity determinations of Y24F and D277N variants of anSMEcpe were similarly carried out with the following modifications: in a final volume of 150 μL either 5 μM Y24F anSMEcpe or 100 μM D277N anSMEcpe enzymes were mixed with 50 mM HEPES, pH 7.5, 150 mM KCl, 1 mM AdoMet, 2 mM sodium dithionite, and 500 μM Kp18Cys. At appropriate times, aliquots were removed, quenched, and analyzed as previously described (26).

III.VI ACKNOWLEDGEMENTS

Data deposition

The atomic coordinates and structure factor discussed in this chapter have been deposited in the Protein Data Bank (PDB ID codes 4K36, 4K37, 4K38, and 4K39).

Author contributions

P.J.G., T.L.G, L.A.S., S.J.B., and C.L.D. designed research; P.J.G. and M.I.M. performed the crystallization and crystal structure determination; T.L.G. provided His_6 and native protein samples for crystallization; L.A.S. generated mutant constructs and performed activity assays; and P.J.G. and C.L.D. wrote the paper.

Table III.1. Data processing and refinement statistics.

	Fe-SAD[1] (His$_6$ construct)	His$_6$, AdoMet-bound	AdoMet-bound	Kp18Cys, AdoMet	Cp18Cys, AdoMet
Beamline	In-house	24-IDC, APS	24-IDE, APS	X29, NSLS	24-IDC, APS
Data Processing					
Wavelength (Å)	1.5418	0.9794	0.9792	1.0750	0.9792
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell Dimensions					
a, b, c (Å)	44.8, 91.9, 90.4	46.6, 92.6, 92.3	44.4, 92.4, 94.1	44.4, 92.1, 91.1	44.0, 91.9, 91.0 91.1
β (°)	91.2	94.0	93.0	91.2	
Resolution (Å)[2]	50.0 – 2.57 (2.66 – 2.57)	50 – 1.62 (1.68 – 1.62)	50.0 – 1.62 (1.68 – 1.62)	50.0 – 1.83 (1.90 – 1.83)	50.0 – 1.79 (1.85 – 1.79)
R_{sym} (%)[2]	6.6 (19.3)	6.1 (57.2)	8.2 (56.0)	11.8 (54.1)	8.3 (37.0)
$<I/\sigma(I)>$[2]	19.3 (5.3)	22.6 (2.5)	16.1 (2.1)	17.2 (2.9)	17.0 (2.6)
Completeness (%)[2]	97.3 (90.3)	98.9 (98.2)	99.8 (98.3)	99.5 (96.4)	98.9 (90.2)
Redundancy[2]	3.9 (3.2)	4.4 (4.2)	3.3 (4.0)	5.9 (5.1)	4.0 (3.5)
Total reflections	162,095	426,267	387,684	380,063	274,984
Refinement					
Resolution (Å)		37 – 1.62	41.5 – 1.62	41.1 – 1.83	41.0 – 1.78
Reflections		97919	95988	64439	68160
R_{work} / R_{free}[3]		18.2 / 19.5	17.2 / 19.9	17.2 / 20.5	18.0 / 21.5
Atoms					
No. of non-hydrogen atoms		6793	7060	6764	6595
Protein		6096	6010	6008	5895
[4Fe4S] / AdoMet / peptide		48 / 54 / -	48 / 54 / -	48 / 54 / 139	48 / 27 / 138
Solvent		595	948	515	487
Average B-factors (Å^2)					
Protein		34.5	16.9	25.0	26.2
[4Fe4S] / AdoMet / peptide		28.2 / 28.3 / -	11.7 / 10.4 / -	19.9 / 21.4 / 36.7	23.1 / 25.3 / 34.6
Solvent		42.3	26.9	31.0	31.6
RMS deviations					
Bond lengths (Å)		0.011	0.013	0.013	0.012
Bond angles (°)		1.40	1.38	1.46	1.35
Ramachandran statistics[4]		93.5%, 6.0%, 0.4%, 0 residues	93.3%, 6.3%, 0.4%, 0 residues	92.4%, 7.3%, 0.3%, 0 residues	94.0%, 5.7%, 0.3%, 0 residues

[1]Data were scaled anomalously.
[2]Highest resolution shell is shown in parentheses.
[3]R-factor $= \Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of 5% of the diffraction data not used in refinement.
[4]Values reported correspond to the number of residues in the most favored, additionally allowed, generously allowed, and disallowed regions, respectively

Figure III.1. anSME reaction. (1) Electron donation to the AdoMet radical cluster initiates homolysis of AdoMet and 5'dA• formation in the presence of bound substrate; (2) substrate radical generation; (3) deprotonation of the substrate Cys sidechain; (4) substrate oxidation; and (5) hydrolysis of the thioaldehyde intermediate yields the formylglycine moiety of the activated sulfatase.
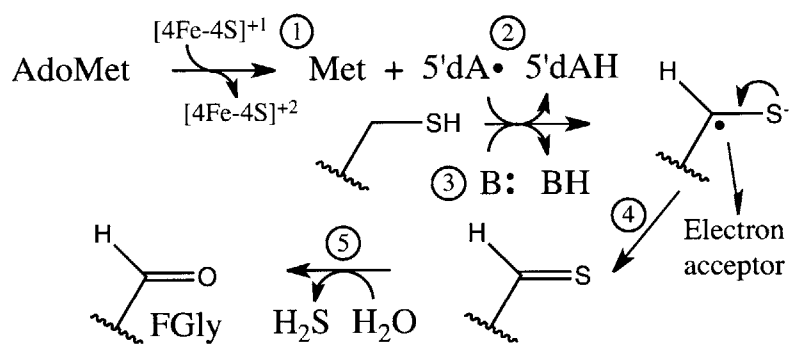
Figure III.2. Structure of anSMEcpe. (*A*) The AdoMet domain (magenta) contains the AdoMet cluster and the $(\beta/\alpha)_6$ partial TIM barrel. The SPASM domain (green) comprises most of the C-terminal segment and houses the remaining two [4Fe-4S] clusters. Two helices, α6a and α6', are not part of either domain and are colored light blue. (*B*) Positions and distances between the three [4Fe-4S] clusters (stick representation with Fe in orange and S in yellow). (*C*) Topology of anSMEcpe. A – AdoMet cluster; I – Auxiliary Cluster I; II – Auxiliary Cluster II. (*D*) The SPASM domain, colored by the level of sequence homology between anSMEcpe and the other 280 members of TIGR04085. Conservation scores were calculated by the ConSurf server (43). Iron ligating cysteines are shown as spheres and labeled by their designation in (*C*). Auxiliary Clusters I and II are shown in stick representation and labeled.
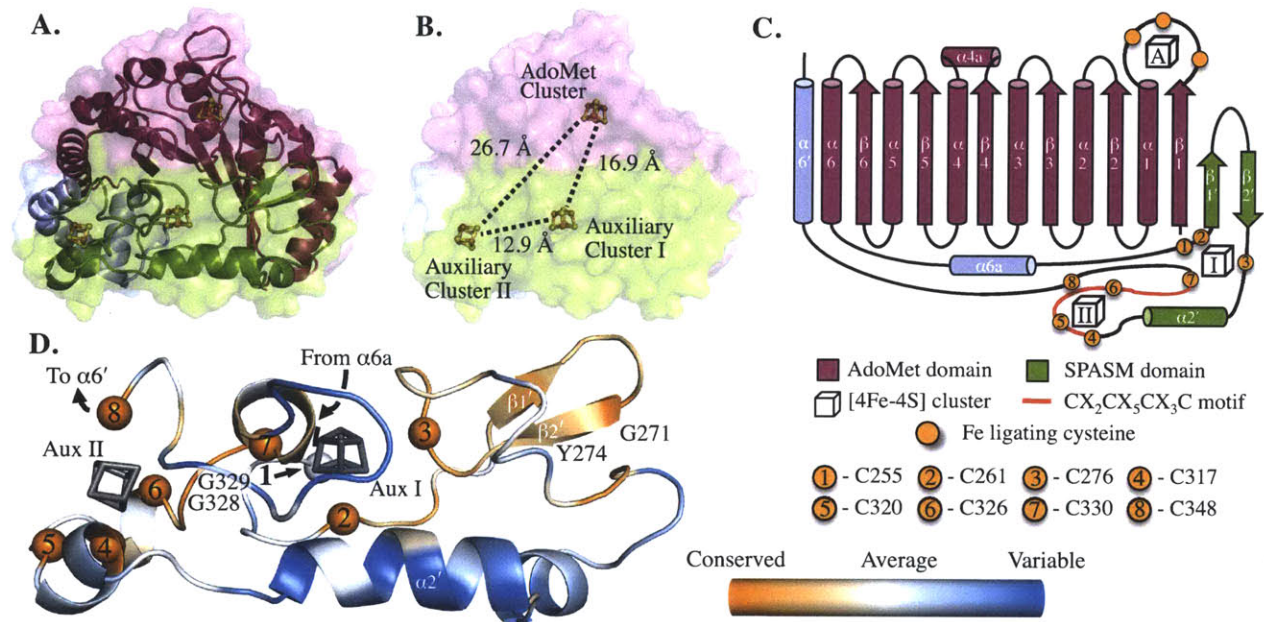
Figure III.3. Conserved elements of AdoMet radical structure. (*A*) $2F_o$-F_c composite omit density contoured to 1.0 σ around AdoMet and the AdoMet cluster. (*B*) The backside of the (β/α)$_6$ barrel, where a conserved patch of residues is found just following β2, a site suggested to have a role in electron donor binding (28, 29). (*C*) AdoMet binding pocket, including the GGE motif (G65, G66, and E67), in which backbone and sidechain oxygen atoms hydrogen bond with the methionine amine group, the ribose motif (S122 and D124), in which polar residues following the β4 strand hydrogen bond with the ribose hydroxyl groups, the 'GXIXGXXE' motif (residues 163–170), where a hydrophobic residue (V165) interacts with the adenine ring of AdoMet and a polar residue in α5 (T170; not shown) stabilizes strand β5 via a backbone hydrogen bond (to the amine of T167 in anSMEcpe; not shown), and the β6 motif, stabilizing and providing backbone hydrogen bonds to the adenine moiety (I192–L195) (28, 29). See Figure III.4.
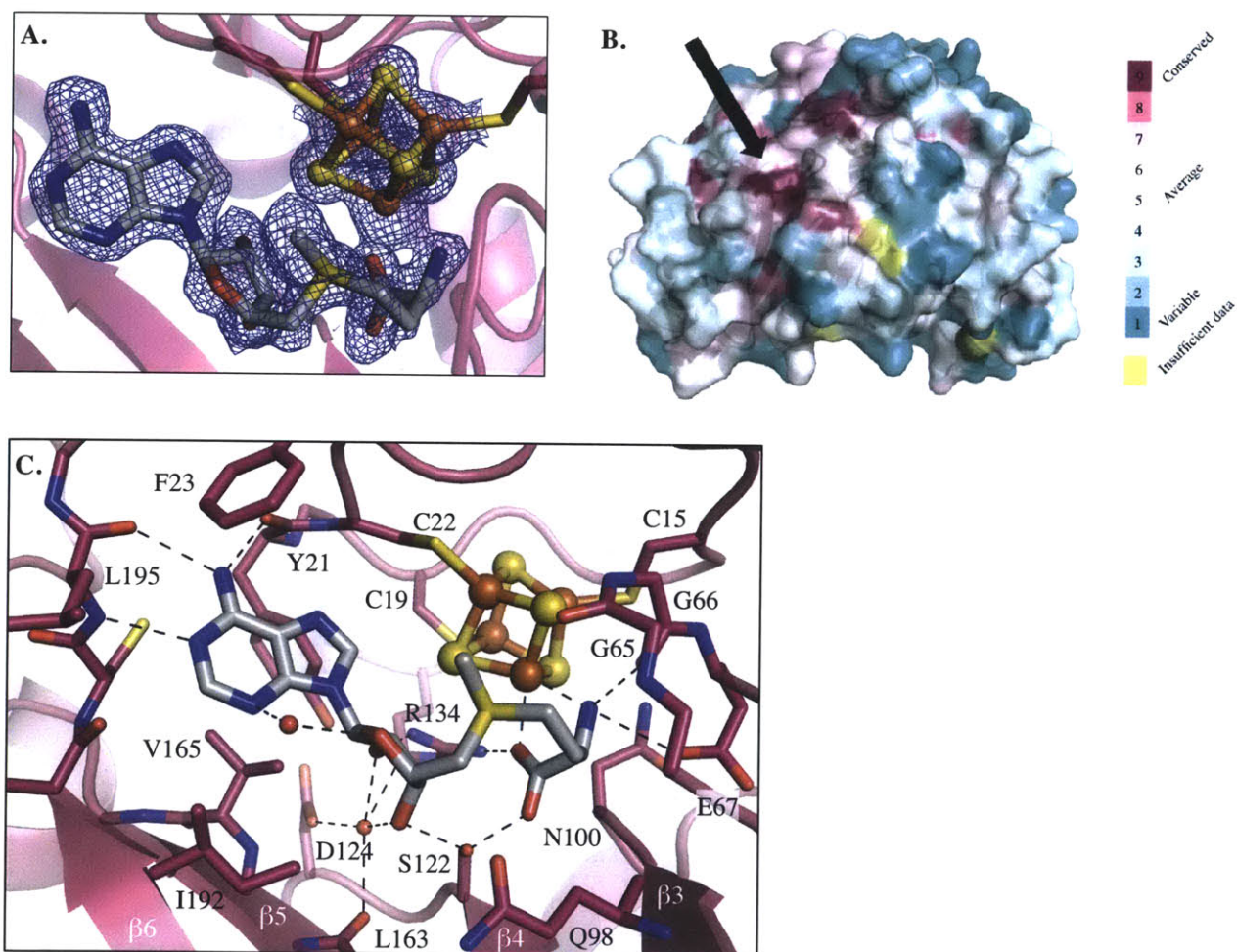
Figure III.4. Alignment of the three biochemically characterized anSMEs. The displayed sequences are for anSMEs from *C. perfringens* (anSMEcpe), *K. pneumoniae* (AtsB), and *B. thetaiotaomicron* (anSMEbt). Highlighted regions in the alignment include the AdoMet cluster binding motif (magenta), Aux I cysteine ligands (yellow), Aux II cysteine ligands (green), $CX_2CX_5CX_3C$ motif (cyan), glutamines involved in substrate binding and orientation (see Figure III.11) and the proposed catalytic residue D277 (blue), and the Arg binding motif (red, see Figure III.9). Conserved AdoMet binding motifs (28, 29) are in grey and labeled.

```
                                                    --β1--
anSMEcpe   --------------------MPPLSLLIKPASSGCNLKCTYCFYHSLSDNRNVKSYGIMR
AtsB       MLNIAALRQQQIPLAAEPRSPVPFHILMKPIGPACNLACRYCYYPQD-----ETPVNKMD
anSMEbt    ----------MKATTYAPFAKPLYVMVKPVGAVCNLACEYCYYLEKANLYKENPKHVMS
                                *:  ::*** .  *** * **:* .      .    *
                                         'CX₃CXΦC' motif
           ~~~~~α1~~~~~~~      --β2--        ~~~~~α2~~~~~~      ---β3-
anSMEcpe   DEVLESMVKRVLNE-ANG--HCSFAFQGGEPTLAGLEFFEKLMELQRKHNYKNLKIYNSL
AtsB       DARLEQFIRRYIAAQPAGAREINFVWQGGEPLLAGLSFYKKALALQARYAPDGVTISNSL
anSMEbt    DELLEKFIDEYINSQTMP--QVLFTWHGGETLMRPLSFYKKAMELQKKYA-RGRTIDNCI
           *  **.:: . :       . *.::*** :  *.*::* : ** ::      .* *.:
                                    GGE motif
           --    ~~~~α3~~~~   --β4-  ~~α4a~~      ~~~~~~α4~~~~~
anSMEcpe   QTNGTLIDESWAKFLSENKFLVGLSMDGPKEIHNLNRKDCCGLDTFSKVERAAELFKKYK
AtsB       QTNGTLINDAWCRLFREHGFIIGLGLEGNEALQDYHRPDKRGRSTWSAALRGIDLLHQHQ
anSMEbt    QTNGTLLTDEWCEFFRENNWLVGVSIDGPQEFHDEYRKNKMGKPSFVKVMQGINLLKKHG
           ******: :  *..:: *. :::*:.::* : :::  *  :  *    :: . :*::::
                                  ribose binding motif
           ---β5-- ~~~~~~~α5~~~~~~~     --β6--
anSMEcpe   VEFNILCVVTSNTARHVNKVYKYFKEKDFKFLQFINCLDPLYEE---------------KG
AtsB       VDFNLLVVVHNEMAAHAAAIYVRLVSLGARYLQFQPLMSEGAA----------------L
anSMEbt    VEWNAMAVVNDFNAEYPLDFYNFFKEIDCHYIQFAPIVERIVSHQDGRHLASLAEGKEGA
           *::* : **  . * :   .* : .   ::**  ::.
             GXIXGXXE motif              β6 structural motif
           ~~~~~~~~~α6~~~~~~~~       ~~~~~~~~~~~~~
anSMEcpe   KYNYSLKPKDYTKFLKNLFDFWYEDFLNGNRVSIRYFDGLLETILLGKSSSCGMNGTCTC
AtsB       REGYQLSADNWGRFMVGIWRQWRKRCDRGRVFVI-NIEQAWAQYFTHTSGSCVHSARCGS
anSMEbt    LADFSVSPEQWGNFLCTIFDEWVKEDVG--KFFIQIFDSTLANWMGEQPGVCTMAKHCGH
           :.:. .:: .*: :: * :         . * ::    :    . *      *
           --hairpin--                 ~~~~~~~~~~~~~~~~~~
anSMEcpe   QFVVESDGSVYPCDFYVLDKWRLGNIQDMTMKELFETNKNHEFIKLSFK-VHEECKKCKW
AtsB       NLVMESDGQLYACDHLINTEHRLGRLDEQTLAAAVDASVQLPFGQQ--KSLRRECQTCSV
anSMEbt    AGVMEFNGDVYSCDHFVFPEYKLGNIYSQTLVEMMHSERQHNFGTMKYQSLPTQCKECDF
           *:* :*.:* **. :  : :**.: . *:  ...: :  *      :  :*: *.
                                    ~~~~~~~~~~~~~~~~~~~
anSMEcpe   FRLCKGGCRRCRDSK-EDSALELNYYCQSYKEFFEYAFPRLINVANNIK-----------
AtsB       KMVCQGGCPAHLN------AAGNNRLCGGYYRFFSDILAPLRPFSRDLNGLKAWRAAFVG
anSMEbt    LFACNGECPKNRFSRTADGEPGLNYLCKGYYQYFQHVAPYMDFMKKELMNQQAPANIMKA
           *:* *         *    * .* .:*.        :   . .::

anSMEcpe   ----------
AtsB       TAHTA-----
anSMEbt    LKDGSLKIEY
```

94

Figure III.5. Structural similarity of anSMEcpe and MoaA. (*A*) anSMEcpe (colored as in Figure III.2) overlays with MoaA (PDB ID 2FB3, grey) (25) with an RMSD of 6.9 Å. (*B*) Both proteins contain the $(\beta/\alpha)_6$ partial TIM barrel common to AdoMet radical proteins. (*C*) The anSMEcpe β-strands are shown in ribbons and Aux I cluster protein ligands shown in sticks (colored as in Figure III.2). AdoMet in black sticks and Kp18Cys peptide in black ribbon, with the target cysteine and anchoring arginine in sticks. The distance from C255 to the hydrogen abstraction site (12.0 Å) is indicated by a dashed line. MoaA clusters and the auxiliary cluster protein ligands are shown in grey sticks. Also in grey sticks is the MoaA substrate, GTP, which is directly ligated to its auxiliary cluster. (*D*) The SPASM domain of anSMEcpe and the C-terminus of MoaA. MoaA contains one cluster in additional to its AdoMet cluster, which overlays well with Aux I in anSMEcpe. Other structural elements superimpose well, including the β1'/β2' hairpin and the α2' helix. The topologies of (*E*) anSMEcpe and (*F*) MoaA are shown in the regions of the auxiliary clusters. In the diagram, Aux I and II are labeled "I" and "II" and the $CX_2CX_5CX_3C$ motif is in red. Iron ligated cysteines are shown as orange circles. In (*F*), a yellow circle represents GTP.

95
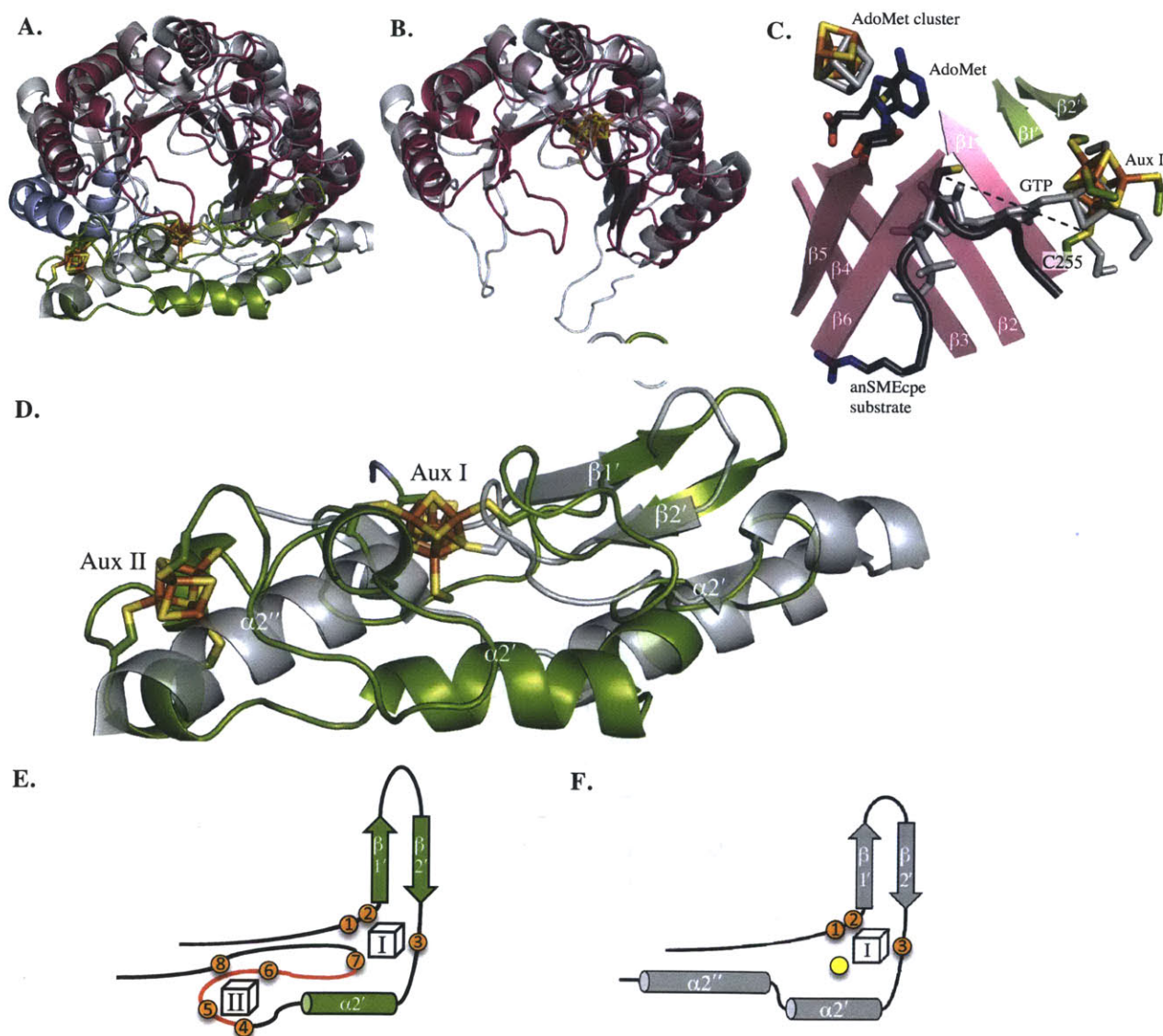
Figure III.6. Peptide substrate sequences and interaction distances with anSMEcpe. (*A*) Comparison of sequences used in crystallography and activity assays (Cp18Cys and Kp18Cys) with sulfatase enzyme sequences from *C. perfringens* (CpArylSulf) and *K. pneumoniae* (AtsA). (*B*) Hydrogen bonding interactions are listed and differ between the two structures in the 4 and 12 positions as indicated.

A.

```
Cp18Cys                Ac-YTAVPSCIPSRASILTGM
CpArylSulf  (46-62)    ...YTAVPSCIASRASILTGM...
Kp18Cys                Ac-YYTSPMCAPARSMLLTGN
AtsA        (66-83)    ...YYTSPMSAPARSMLLTGN...
                        1    5     10   15
```

B.

Peptide H-bond atom	anSMEcpe H-bond atom	Distance
4 – backbone NH	Ser253 – backbone CO	2.9 – 3.0 Å
4 – Kp18Cys only; Ser Oγ	Ser253 – backbone CO	3.5 – 3.6 Å
4 – Kp18Cys only; Ser Oγ	Arg238 – NH$_2$(a)	2.3 – 2.5 Å
4 – backbone CO	Cys255 – backbone NH	2.8 – 3.0 Å
4 – backbone CO	Arg238 – NH$_2$(a)	2.8 – 2.9 Å
5 – backbone CO	Arg238 – NH$_2$(a)	2.8 – 3.5 Å
5 – backbone CO	Arg238 – NH$_2$(b)	3.3 Å
6 – backbone CO	Arg238 – NH$_2$(b)	3.2 – 3.5 Å
7 (Cys) – backbone CO	Gln64 – NH$_2$	3.2 – 3.6 Å
8 – backbone CO	Arg238 – NH$_2$(b)	3.0 – 3.4 Å
8 – backbone CO	Gln190 – NH$_2$	2.9 – 3.0 Å
10 – backbone NH	Gln190 – CO	2.9 – 3.0 Å
10 – backbone CO	Asn161 – NH$_2$	2.7 – 2.8 Å
11(Arg) – Nε	Asn161 – CO	2.7 – 2.8 Å
11(Arg) – NH$_2$(a)	Asn161 – CO	3.1 – 3.2 Å
11(Arg) – NH$_2$(a)	Phe190 – backbone CO	2.8 – 3.1 Å
11(Arg) – NH$_2$(a)	Glu159 – COO	3.4 – 3.5 Å
11(Arg) – NH$_2$(b)	Glu159 – COO	2.7 – 2.8 Å
12 – Kp18Cys only; Ser Oγ	Glu245 – COO	3.0 Å
13 – backbone NH	Glu245 – COO	3.4 Å

* Intrapeptide H-bond: 6-CO – 8-NH; 3.3 – 3.4 Å
* Does not include multiple water-meditated interactions

Figure III.7. anSMEcpe and peptide substrate interactions. (*A*) Kp18Cys peptide and Auxiliary Cluster I simulated annealing $2F_o$-F_c composite omit density, contoured at 1.0 σ. (*B*) The Cp18Cys and Kp18Cys bound structures overlaid. (*C*) Kp18Cys hydrogen bonding network. Only the 4 and 12 positions differ in hydrogen bonding network due to a serine in Kp18Cys and a valine in Cp18Cys in position 4 and a serine in Kp18Cys and an alanine in Cp18Cys in position 12 (see Figure III.6). Coloring similar to Figure III.2.

97
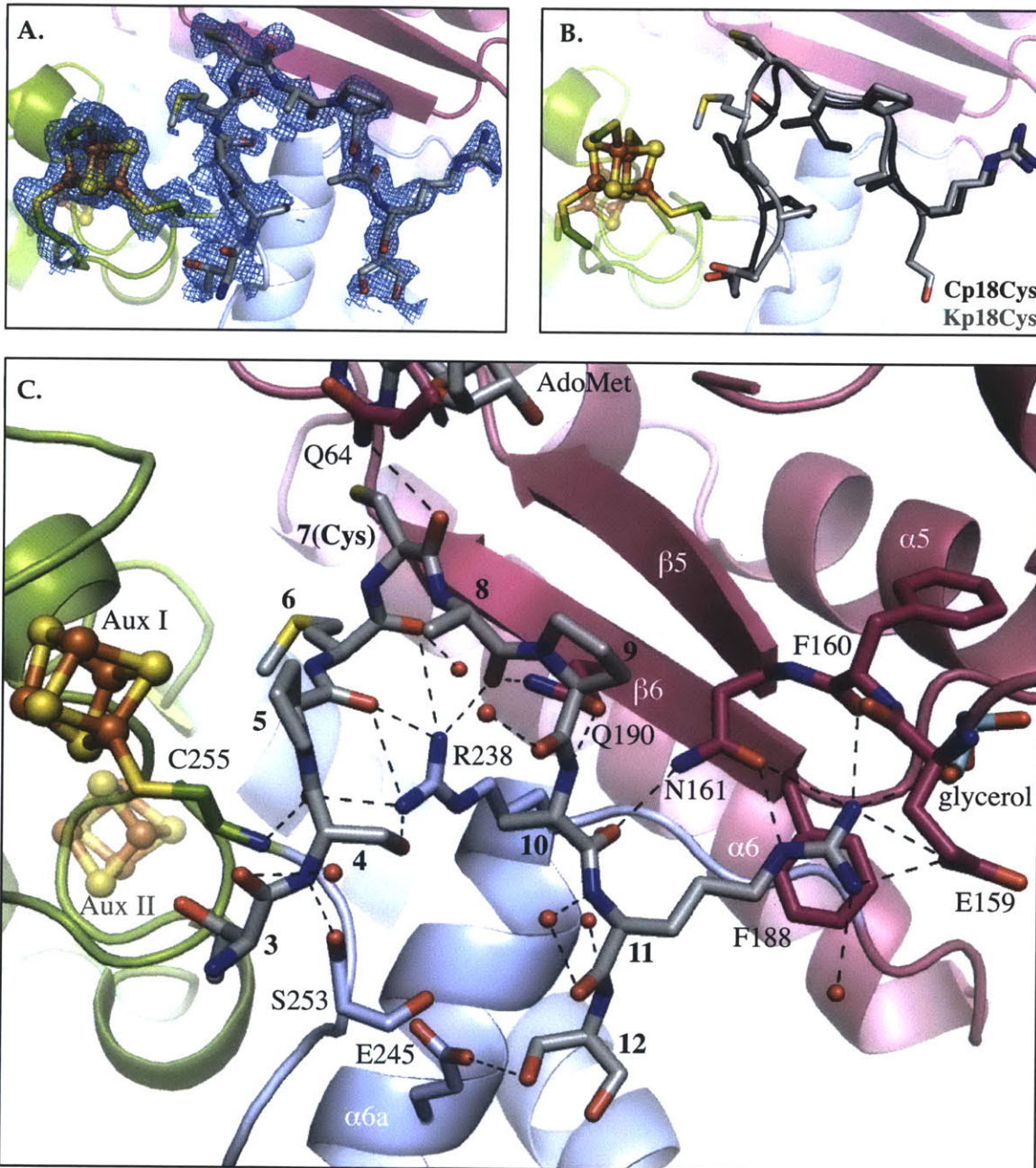
Figure III.8. Substrate peptide binding. (*A*) Cp18Cys (black) and Kp18Cys (grey) enter and exit the active site of anSMEcpe via the underside of the barrel. The Cβ carbon is 8.6 and 8.9 Å from Aux I and the AdoMet cluster, respectively. AdoMet and auxiliary clusters are shown in sticks with carbons in grey, oxygens in red, nitrogens in blue, sulfurs in yellow, and irons in orange. anSMEcpe β strands and SPASM domain are shown in ribbons and colored as in Figure III.2. (*B*) Substrate peptides bound to anSMEcpe and (*C*) bound to FGE (PDB 2AIJ, blue) (7) were overlaid by the five residues encompassing the conserved sulfatase motif in each system and are shown in the same orientation.
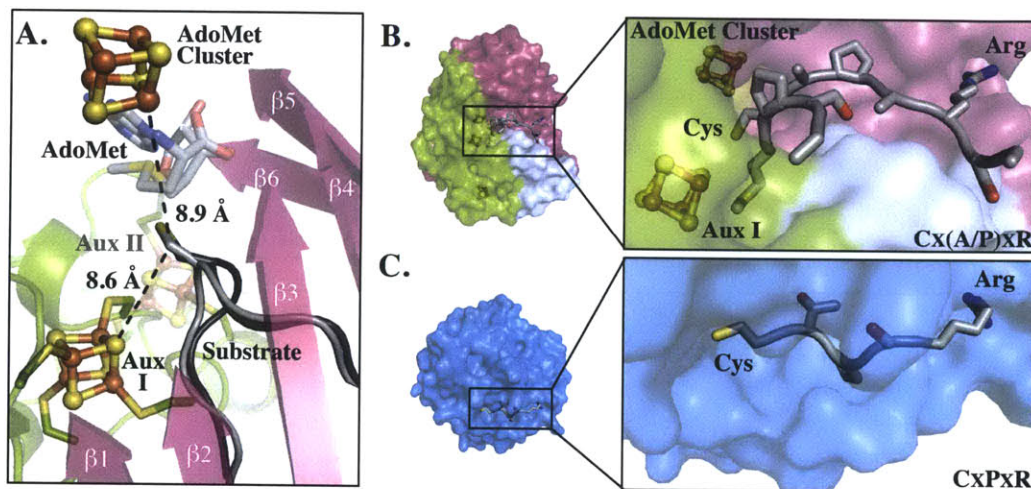
Figure III.9. Comparison of peptidyl-substrate sequence specificity in anSME and FGE. The arginine in the conserved CX(P/A)XR motif in (*A*) anSMEcpe and (*B*) FGE (PDB ID 2AIJ) (7). Both systems use a phenylalanine group for π stacking and an acidic residue for salt bridge interactions. See Figure III.4 for the sequence alignment of the anSME arginine binding site. Colored as in Figure III.8. (*C*) At position 6 of the bound-peptide substrates Cp18Cys and Kp18Cys in anSMEcpe, a large solvated internal enzyme cavity is able to accommodate either the sidechain of Ala (Cp18Cys, black) or of Met (Kp18Cys, grey). (*D*) At position 8 of the bound-peptide in anSMEcpe, a cavity is created by the presence of an Ala on β2 (A62, in sticks). Here, I8 (Cp18Cys, black) packs well against A62, whereas when A8 (Kp18Cys, grey) is accommodated, a space is left between the two Ala sidechains. In the native *K. pneumoniae* anSME, AtsB, A62 is replaced by valine, which would remove this gap between enzyme and the Kp18Cys target peptide.
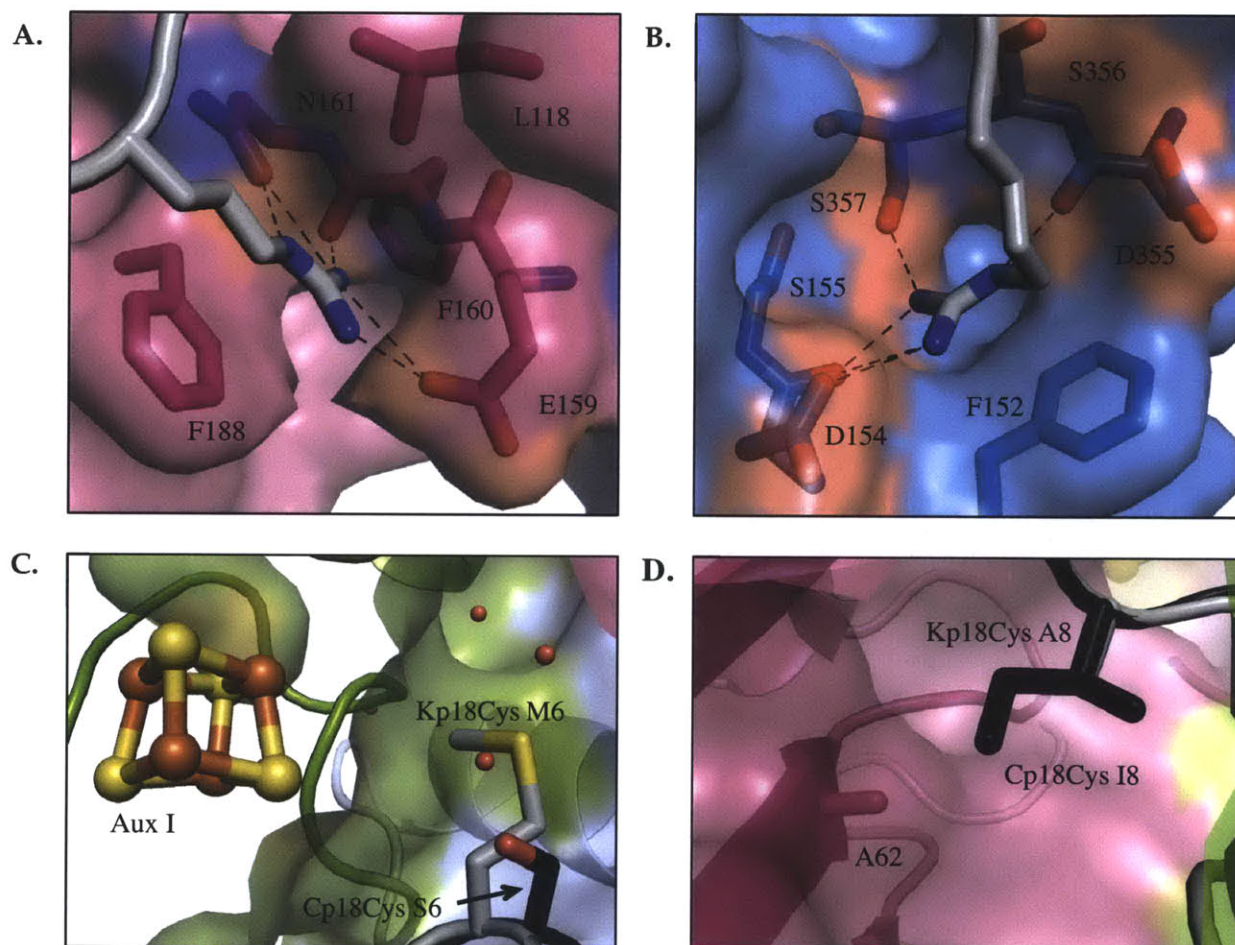
Figure III.10. anSME active site. (*A*) The active site of anSMEcpe. Sticks are displayed for AdoMet, target cysteine, and residues within 5 Å of the substrate cysteine Sγ. Distances as follows; AdoMet 5′C – cysteine Cβ, 4.1 Å; Y24 – Sγ, 4.7 Å; D277 – Sγ, 4.6 Å; Q64 – Sγ, 3.3 Å. (*B*) FGly and 5′dA production for the Y24F and D277N mutants. Displayed product formation is per μmol enzyme. *Wild-type data from (26).
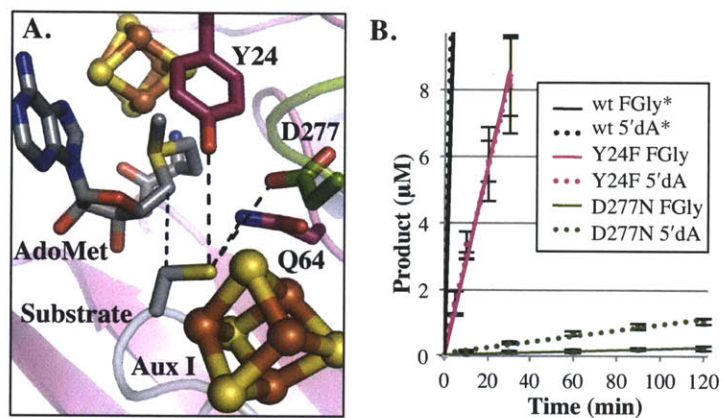
Figure III.11 Mobile Glns in the anSMEcpe structures. (*A*) The Kp18Cys active site, Position 3 as described in the Results, in lighter and darker colors are the Cp18Cys structure (Position 2) and the AdoMet structure (Position 1), respectively. Multiple conformations of D277 are observed in the AdoMet and Cp18Cys structures. Kp18Cys H-bonding network is in black dashes and substrate deprotonation distance is in red dashes. (*B*) Mobile Glns in the Kp18Cys structure (Position 3 as the Results) shown with $2F_o$-F_c electron density, contoured at 1.0 σ.
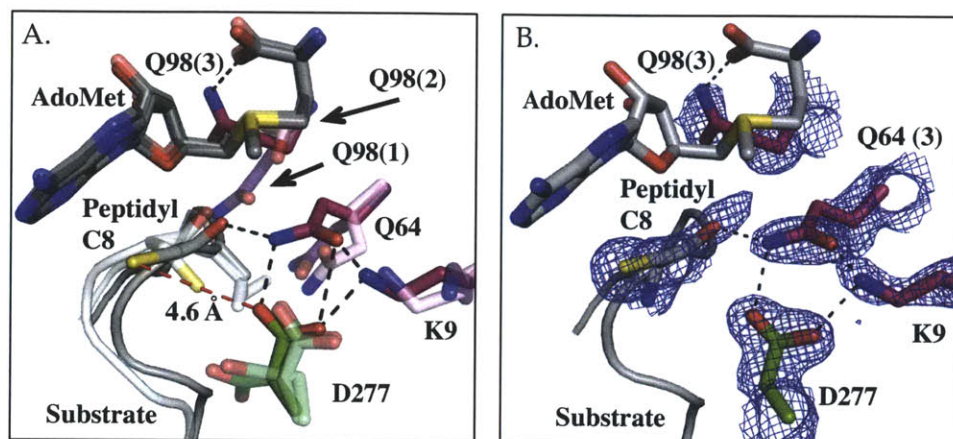
Figure III.12. Changes in residue distance to bulk solvent following peptide binding. anSMEcpe is colored by depth from bulk solvent in the (*A*) peptide free and (*B*) Kp18Cys bound (peptide in thick ribbon) structures. The C-terminal domain and interior β sheet of the AdoMet domain are shown in ribbon representation, with AdoMet and [4Fe4S] clusters in sticks. PDBs generated by the DEPTH server (44).
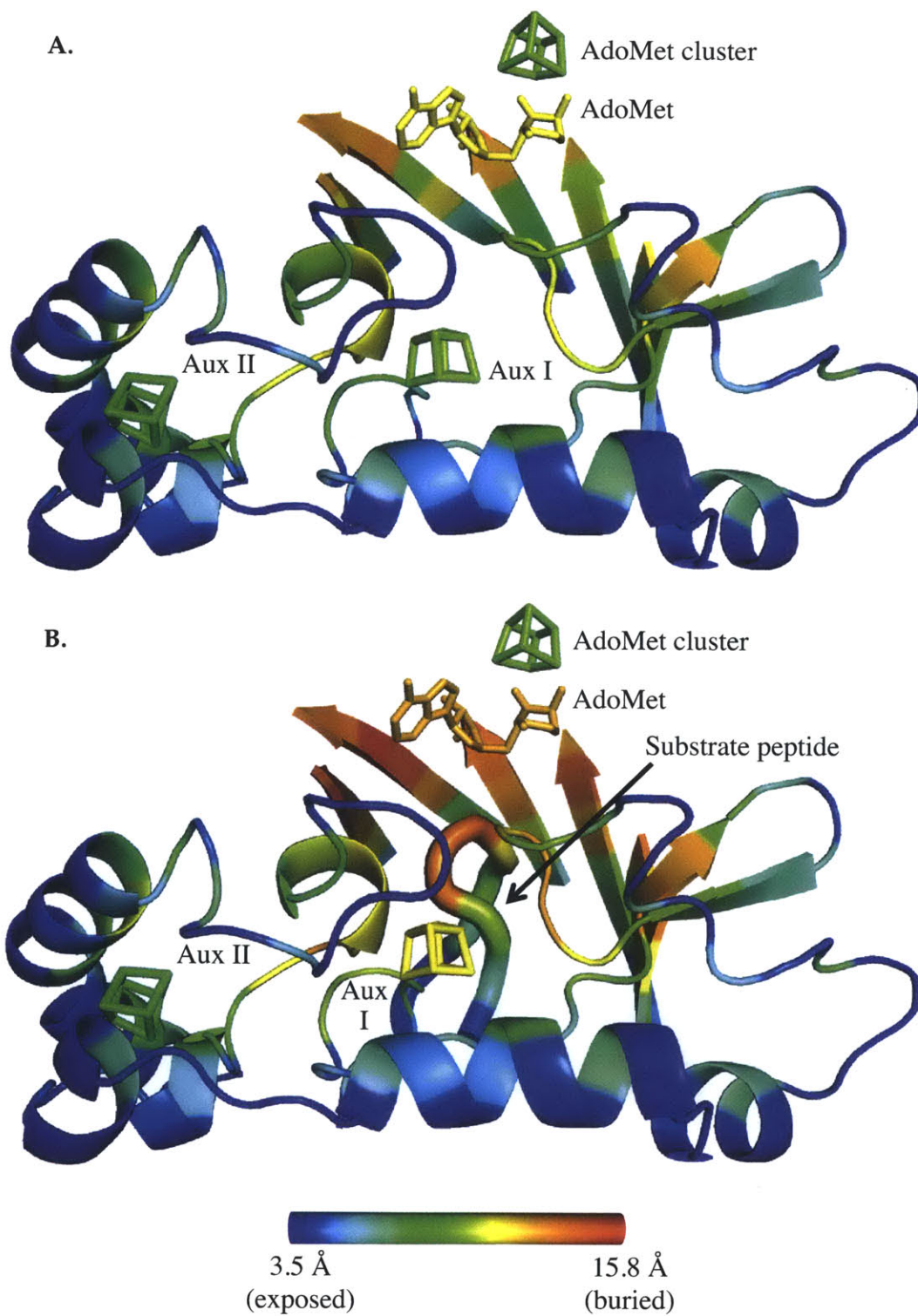
Figure III.13. Proposed mechanism of anSMEcpe. The initiation of turnover in anSMEcpe is the reductive homolysis of AdoMet into 5'dA• and methionine (Figure III.1) and the deprotonation of the sidechain by the putative base D277. The orientation of substrate implies that the Cβ pro-*S* hydrogen is then abstracted, resulting in the activated radical thiolate. Upon collapse of the thiolate to the thiolaldehyde, the unpaired electron is passed to Aux I, where it can transfer to Aux II for subsequent reduction of a terminal electron acceptor. The thiolaldehyde is hydrolyzed to the active FGly moiety.
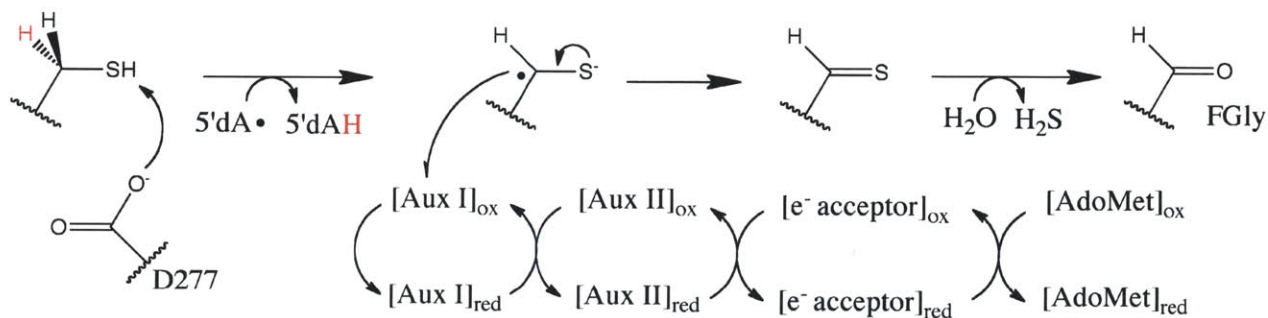
Figure III.14. The AdoMet radical superfamily. A modification of the Enzyme Function Initiative's Radical SAM Superfamily Workshop materials (http://enzymefunction.org/resources/workshops). The entire superfamily (nodes are enzymes with 40% sequence identity) is shown using a threshold cut-off of an E-Value of 1 x 10^{-22}. Figure generated using Cytoscape (45).

○ SPASM Domain Proteins (TIGR04085); Reactions Unknown
○ Dehydrogenases
○ Complex Rearrangement
○ Reactions Unknown
● B12-like Rearrangement
○ Glycyl Radical Activating enzymes
● Metallocofactor Assembly
● Methylation Reactions (B12 Dependent)
● Methylation Reactions (non-B12)
● Sulfur Insertion

△ Experimental Evidence
◇ Predicted

⇨ Structure Available
⇨ This work

Notable Protein Nodes

1. Cys-type anSMEs (anSMEcpe)
2. Ser-type anSMEs (AtsB)
3. MtfC
4. KwcM
5. MoaA
6. BtrN
7. AlbA
8. PqqE
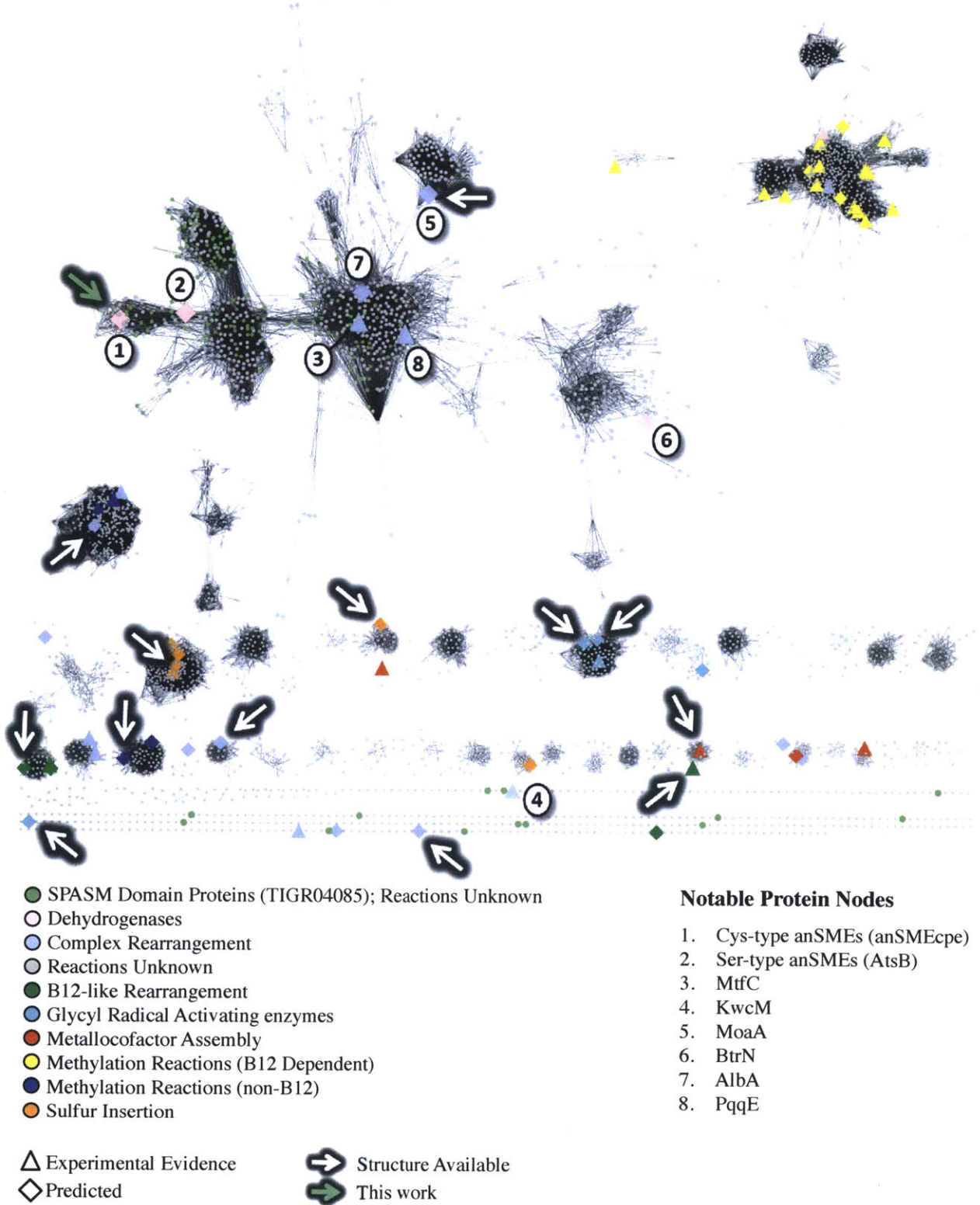
III.VII   REFERENCES

1. Schmidt B, Selmer T, Ingendoh A, & von Figura K (1995) A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* 82(2):271-278.

2. Ghosh D (2007) Human sulfatases: a structural perspective to catalysis. *Cell Mol Life Sci* 64(15):2013-2022.

3. Bojarova P & Williams SJ (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Curr Opin Chem Biol* 12(5):573-581.

4. Dierks T, *et al.* (2003) Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C-alpha-formylglycine generating enzyme. *Cell* 113(4):435-444.

5. Benjdia A, Martens EC, Gordon JI, & Berteau O (2011) Sulfatases and a radical AdoMet enzyme are key for mucosal glycan foraging and fitness of a prominent human gut Bacteroides. *J Biol Chem* 286(29):25973-25982.

6. Dierks T, *et al.* (2005) Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme. *Cell* 121(4):541-552.

7. Roeser D, *et al.* (2006) A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. *Proc Natl Acad Sci USA* 103(1):81-86.

8. Fang Q, Peng J, & Dierks T (2004) Post-translational formylglycine modification of bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. *J Biol Chem* 279(15):14570-14578.

9. Benjdia A, *et al.* (2007) Anaerobic sulfatase-maturating enzymes: radical SAM enzymes able to catalyze in vitro sulfatase post-translational modification. *J Am Chem Soc* 129(12):3462-3463.

10. Grove TL, Lee KH, St Clair J, Krebs C, & Booker SJ (2008) In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry* 47(28):7523-7538.

11. Berteau O, Guillot A, Benjdia A, & Rabot S (2006) A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. *J Biol Chem* 281(32):22464-22470.

12. Rabuka D, Rush JS, deHart GW, Wu P, & Bertozzi CR (2012) Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nature Protocols* 7(6):1052-1067.

13. Frey PA, Hegeman AD, & Ruzicka FJ (2008) The Radical SAM Superfamily. *Crit Rev Biochem Mol Biol* 43(1):63-88.

14. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, & Miller NE (2001) 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):1097-1106.

15. Hiscox MJ, Driesener RC, & Roach PL (2012) Enzyme catalyzed formation of radicals from S-adenosylmethionine and inhibition of enzyme activity by the cleavage products. *Biochim Biophys Acta* 1824(11):1165-1177.

16. Walsby CJ, Ortillo D, Broderick WE, Broderick JB, & Hoffman BM (2002) 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):11270-11271.

17.	Benjdia A, et al. (2010) Anaerobic sulfatase-maturating enzyme - a mechanistic link with glycyl radical-activating enzymes? *FEBS J* 277(8):1906-1920.

18.	Yokoyama K, Ohmori D, Kudo F, & Eguchi T (2008) Mechanistic study on the reaction of a radical SAM dehydrogenase BtrN by electron paramagnetic resonance spectroscopy. *Biochemistry* 47(34):8950-8960.

19.	Yokoyama K, Numakura M, Kudo F, Ohmori D, & Eguchi T (2007) Characterization and mechanistic study of a radical SAM dehydrogenase in the biosynthesis of butirosin. *J Am Chem Soc* 129(49):15147-15155.

20.	Lanz ND & Booker SJ (2012) Identification and function of auxiliary iron-sulfur clusters in radical SAM enzymes. *Biochim Biophys Acta* 1824(11):1196-1212.

21.	Grove TL, Ahlum JH, Sharma P, Krebs C, & Booker SJ (2010) A consensus mechanism for Radical SAM-dependent dehydrogenation? BtrN contains two [4Fe-4S] clusters. *Biochemistry* 49(18):3783-3785.

22.	Haft DH & Basu MK (2011) Biological systems discovery in silico: Radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol* 193(11):2745-2755.

23.	Haft DH (2011) Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* 12:21.

24.	Lees NS, et al. (2009) ENDOR spectroscopy shows that guanine N1 binds to [4Fe-4S] cluster II of the S-adenosylmethionine-dependent enzyme MoaA: Mechanistic implications. *J Am Chem Soc* 131(26):9184-9185.

25.	Hanzelmann P & Schindelin H (2006) 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 USA* 103(18):6829-6834.

26.	Grove TL, et al. (2013) Further characterization of Cys-type and Ser-type anaerobic sulfatase maturating enzymes suggests a commonality in the mechanism of catalysis. *Biochemistry* 52(17):2874-2887.

27.	Benjdia A, et al. (2008) Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. *J Biol Chem* 283(26):17815-17826.

28.	Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.

29.	Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.

30.	Layer G, Moser J, Heinz DW, Jahn D, & Schubert WD (2003) Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J* 22(23):6214-6224.

31.	Nicolet Y, Amara P, Mouesca JM, & Fontecilla-Camps JC (2009) Unexpected electron transfer mechanism upon AdoMet cleavage in radical SAM proteins. *Proc Natl Acad Sci USA* 106(35):14867-14871.

32.	Vey JL, et al. (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

33.	Hanzelmann P & Schindelin H (2004) Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc Natl Acad Sci USA* 101(35):12870-12875.

34. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774-797.

35. Moser CC, Anderson JL, & Dutton PL (2010) Guidelines for tunneling in enzymes. *Biochim Biophys Acta* 1797(9):1573-1586.

36. Grove TL, *et al.* (2011) A radically different mechanism for S-adenosylmethionine-dependent methyltransferases. *Science* 332(6029):604-607.

37. Szu PH, Ruszczycky MW, Choi SH, Yan F, & Liu HW (2009) Characterization and mechanistic studies of Desll: A radical S-adenosyl-L-methionine enzyme involved in the biosynthesis of TDP-D-Desosamine. *J Am Chem Soc* 131(39):14030-14042.

38. Mehta AP, *et al.* (2013) Catalysis of a new ribose carbon-insertion reaction by the molybdenum cofactor biosynthetic enzyme MoaA. *Biochemistry* 52(7):1134-1136.

39. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

40. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

41. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.

42. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213-221.

43. Ashkenazy H, Erez E, Martz E, Pupko T, & Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* 38:W529-W533.

44. Tan KP, Varadarajan R, & Madhusudhan MS (2011) DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Research* 39:W242-W248.

45. Smoot ME, Ono K, Ruscheinski J, Wang PL, & Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27(3):431-432.

Chapter 4.

Structural Analysis of BtrN, a Radical DOIA Dehydrogenase, Uncovers an Abundant AdoMet Radical Auxiliary [4Fe-4S] Cluster Binding Domain

A version of this chapter will be submitted for publication.

Authors:
Peter J. Goldman,[1] Tyler L. Grove,[4] Squire J. Booker,[4,5] and Catherine L. Drennan[1,2,3]

[1]Department of Chemistry, [2]Department of Biology, and the [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA
[4]Department of Chemistry and [5]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802 USA

IV.I SUMMARY

BtrN, a 2-deoxy-*scyllo*-inosamine (DOIA) dehydrogenase, is a member of a recently described subclass of the *S*-adenosyl-L-methionine (AdoMet) radical superfamily of enzymes, the AdoMet radical dehydrogenases. Unlike traditionally NAD-dependent DOIA dehydrogenases in other 2-deoxystreptamine (DOS) biosynthetic pathways, the organism *Bacillus circulans* uses a molecule of AdoMet and two [4Fe-4S] clusters to catalyze this step. One cluster is for the reductive homolysis of AdoMet, common to all members of the AdoMet radical superfamily, while the function of the other cluster was unknown. Here, we use structural methods to show that contrary to the current hypothesis, BtrN does not use its second, auxiliary [4Fe-4S] cluster to bind substrate, suggesting it instead plays a role in the oxidation of a substrate intermediate. We find that the protein is responsible for binding DOIA and contributing the putative base in the reaction, R152. In addition, residues at the C-terminus of the protein bind substrate and seal the active site from solvent. The finding that DOIA does not directly ligate the auxiliary cluster in BtrN agrees with recent studies on another AdoMet radical dehydrogenase, the anaerobic sulfatase maturing enzyme (anSME). The auxiliary cluster binding domains of BtrN and anSME are very similar, despite differences in [4Fe-4S] content (anSME contains three [4Fe-4S] clusters). The structural homology between BtrN, anSME, and another AdoMet radical enzyme, MoaA, involved in molybdenum cofactor biosynthesis, suggests that their shared auxiliary cluster binding architecture will be a staple in the AdoMet radical superfamily, common to ~ 30% of the AdoMet radical reactions.
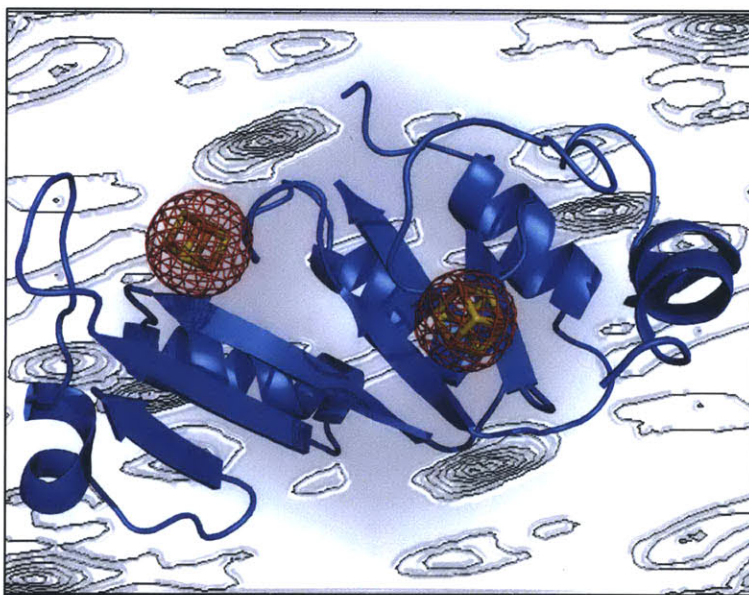
IV.II INTRODUCTION

Due to mounting antibiotic resistance, the discovery and/or modification of antibiotic compounds to fight bacterial infection is a vital task of our time (1). Understanding antibiotic biosynthesis is an important first step in being able to engineer a new wave of therapeutic molecules. Aminoglycosides are a class of antibiotics that inhibit protein synthesis by interacting with the 30S subunit of the bacterial ribosome and have had considerable success in the clinical setting (2). Most FDA approved aminoglycosides, including neomycin, gentamicin, and kanamycin, share a common glucose-6-phosphate derived 2-deoxystreptamine (DOS) structural core (blue in Figure IV.1) (3). While the majority of aminoglycoside-producing bacteria employ similar reaction sequences to biosynthesize this DOS core, including the penultimate two-electron oxidation of 2-deoxy-*scyllo*-inosamine (DOIA) to amino-dideoxy-*scyllo*-inosose (amino-DOI) (Figure IV.1), identification of the enzymes responsible has not always been straightforward. For example, the *btrE* gene product in the butirosin B producing *Bacillus circulans* was proposed to be an NAD-dependent enzyme responsible for the generation of amino-DOI, as in other aminoglycoside pathways (4). This hypothesis proved incorrect upon further sequence and biochemical analysis (5, 6). Instead, unlike other DOS pathways, Yokoyama *et al.* found that production of amino-DOI required another enzyme, BtrN, an *S*-adenosyl-L-methionine (AdoMet, SAM) radical dehydrogenase (5). Here we report structures of catalytic and noncatalytic forms of BtrN, providing a structural basis for the unique reaction it performs.

The AdoMet radical enzyme family catalyzes a diverse array of radical based reactions, including sulfur insertions, complex chemical transformations and rearrangements, DNA and RNA modifications, and, in the case of BtrN, dehydrogenation (7). In addition to DesII, a deaminase involved in the biosynthesis of TDP-D-desosamine, BtrN is the second characterized AdoMet radical carbohydrate-tailoring enzyme. BtrN is also one of two members of the dehydrogenase subfamily of AdoMet radical enzymes that has been characterized biochemically. The other is the anaerobic sulfatase maturating enzyme (anSME) family, which perform a two electron oxidation of a serine or cysteine to generate a formyl glycine residue on a sulfatase protein substrate (Figure IV.1) (8, 9). BtrN and the anSMEs are thought to use identical reaction mechanisms (10). BtrN, DesII, and anSME, like other AdoMet radical family members, require a [4Fe-4S] cluster, a molecule of AdoMet, and a reducing equivalent to initiate turnover (5). This required [4Fe-4S] cluster is ligated by three cysteines of a CX_3CX_2C motif (11, 12), leaving the fourth iron available to bind AdoMet (13). Electron transfer from the [4Fe-4S] cluster to AdoMet causes homolysis of a carbon-sulfur bond in AdoMet, forming methionine

and a 5'-deoxyadenosyl radical (5'dA•). This radical species subsequently abstracts a hydrogen atom from substrate, resulting in 5'-deoxyadenosine (5'dAH) and a substrate radical. While similarities are observed in how AdoMet radical enzymes bind the cluster and AdoMet, leading to a description of an 'AdoMet radical core', structures vary in substrate binding regions which serves to differentiate between the chemistry performed.

Further biochemical analysis of BtrN showed that it is the hydrogen from the C3 position of DOIA that is transferred to 5'dA• during catalysis (Figure IV.1) (5), resulting in 5'dAH and a radical intermediate localized at the C3 position (14). Following this hydrogen abstraction, deprotonation and one electron oxidation of substrate yield amino-DOI product. Thus catalysis requires both a proton and an electron acceptor. Interestingly, both BtrN and anSMEs house [4Fe-4S] clusters in addition to the one required for AdoMet homolysis (10). BtrN binds one additional cluster (15), while the anSMEs bind two (9, 16), suggesting that at least one auxiliary (Aux) cluster is required for AdoMet radical dehydrogenase chemistry (15). Direct ligation of substrate to an Aux cluster became an attractive hypothesis for the role of these clusters, as it would aid in both deprotonation and oxidation of the substrate intermediate (9, 15). However, the recently determined structure of anSME from *Clostridium perfringens* (anSMEcpe) solved with peptide substrate bound showed that these Aux clusters are fully ligated by cysteine residues from the protein and do not play a role in substrate binding (17).

In addition to this mechanistic insight, the anSMEcpe structure provided the first characterization of the SPASM domain. The SPASM proteins are an AdoMet radical enzyme subfamily thought to use two Aux clusters in the maturation of ribosomally translated peptides (18). The subclass is named for its biochemically characterized founding members, AlbA (19), PqqE (20), anSMEs (9, 16), and MtfC (21), involved in subtilosin A, pyrroloquinoline quinone, anaerobic sulfatase, and mycofacticin maturation, respectively. Based on work done by Benjdia, *et al.* (16), the SPASM subfamily was defined by Haft and Basu as 281 AdoMet radical enzymes with a C-terminal 'CX$_{9-15}$GX$_4$C—gap—CX$_2$CX$_5$CX$_3$C—gap—C' motif (18). Structural features of the first half of the SPASM domain are very similar to those seen earlier in the Aux cluster binding domain of the AdoMet radical enzyme MoaA, which catalyzes a dramatic rearrangement of GTP (22-24). These conserved features include a beta hairpin surrounded by iron ligating cysteine positions and followed by a helical region. Due to its partial-SPASM makeup, this substructure was named a 'twitch' domain. As BtrN contains these sequence elements (CX$_{9-15}$GX$_4$CX$_n$), it was hypothesized to contain its Aux cluster in a twitch domain similar to MoaA (17). Interestingly, the Aux cluster in MoaA has an available iron ligation site which, unlike in anSMEcpe, is used to bind substrate (25, 26).

In this manuscript, we use structural methods to address the uncertainty in the role of the Aux cluster in BtrN and to expand our understanding of the AdoMet radical dehydrogenase subfamily. To this end, we have solved crystal structures of BtrN in two forms, a noncatalytic, OPEN, conformation, which does not support substrate binding, and a catalytic, CLOSED, form, in which both DOIA and AdoMet are bound. Using these structures, we conclusively address the hypothesis of direct substrate ligation to the Aux cluster in BtrN, which, as in anSMEcpe, is inaccurate. The [4Fe-4S] cluster architecture and substrate binding patterns in BtrN show surprising similarities to the anSMEcpe system. In addition, we find that the AdoMet radical fold of BtrN lacks 25% of the secondary structure thought to be strictly necessary for AdoMet radical chemistry. Despite its unique, abridged AdoMet radical fold, BtrN utilizes all previously described AdoMet binding interactions, ensuring a relatively conserved primary coordination sphere for the cofactor.

IV.III RESULTS

Anomalous signal from a SeMet derivatized C-terminally His_6 tagged BtrN dataset collected at the selenium edge was used to phase an initial structure of the enzyme (Table IV.1). This structure, solved to 2.02 Å resolution, represents an OPEN conformation of BtrN as the C-terminal region is extended from the protein core, leaving the active site highly solvent exposed (Figure IV.2). In this OPEN conformation, residues in the α4 helix (121-134), a loop following the AdoMet cluster binding loop (28-32), and a linker region (146-161) connecting the N-terminal AdoMet domain to the C-terminal auxiliary cluster domain are disordered and not included in the model. In addition, electron density for BtrN substrates AdoMet and DOIA, which were included in the crystallization conditions, is not observed in this OPEN structure. To obtain a CLOSED structure of BtrN, a native (non-SeMet) N-terminally His_6 tagged construct was used. A structure of this construct, in which the entire BtrN fold was apparent, was solved to 1.56 Å resolution. In this model, a CLOSED conformation of the protein is observed, with the C-terminus capping the enzyme's active site. This CLOSED structure includes the disordered regions not present in the OPEN structure and clear electron density for both BtrN substrates. Both the OPEN and CLOSED structures are constituted with two [4Fe-4S] clusters and in each case, the Aux cluster is fully protein ligated. The overall fold of BtrN includes a partial AdoMet radical fold (residues 1-150), an auxiliary cluster binding motif (residues 169-235), and two loop regions – a linker joining the AdoMet and auxiliary cluster domains (residues 151-168) and the C-terminal cap (residues 236-250) (Figure IV.3).

113

The typical AdoMet radical fold consists of a conserved β_6/α_6 partial TIM barrel fold with a [4Fe-4S] cluster at the top of the barrel in a loop following $\beta 1$ (27, 28). BtrN initiates in the same manner with C16, C20, and C23 providing three ligands to its AdoMet radical cluster. The full β_6/α_6 is not conserved in BtrN, which instead has a β_5/α_4 structure before ending the AdoMet radical domain (Figure IV.3). Despite this difference in fold, interactions provided by all four AdoMet binding motifs common to AdoMet radical proteins, including the 'GGE' motif, the ribose motif, the 'GXIXGXXE' motif, and the $\beta 6$ motif (28), are for the most part conserved in BtrN (Figure IV.4). In addition, the fold contains a basic residue, H117, that interacts with the carboxyl group of AdoMet, as seen in HemN, HydE, PylB, and anSMEcpe (17, 29-31), and a hydrogen bond to N6 of the adenine from the carbonyl of Y22, the hydrophobic residue of the [4Fe-4S] cluster binding $CX_3CX\Phi C$ motif, common to nearly all AdoMet radical members (Figure IV.5) (28). The Y22 interaction is disrupted in the OPEN structure, as the conformation of W21 is altered due to crystal packing, distorting the AdoMet cluster binding loop and possibly contributing to the inability of the OPEN structure to bind AdoMet (Figure IV.2).

While BtrN contains all AdoMet binding motifs, the 'GXIXGXXE' and $\beta 6$ motifs are altered because of the abridged nature of the enzyme's AdoMet radical fold. The 'GXIXGXXE' motif provides a hydrophobic contact at the end of $\beta 5$ to the adenine ring of AdoMet and is stabilized by a backbone interaction with a polar residue in $\alpha 5$. BtrN has no $\alpha 5$ and its $\beta 5$ strand terminates before reaching the adenine moiety (Figure IV.6). Following the $\beta 5$ strand in BtrN, however, a loop containing L147 provides a similar hydrophobic interaction to the adenine as found in the 'GXIXGXXE' motif of other AdoMet radical enzymes (27, 28). Instead of being stabilized by a polar residue from $\alpha 5$, this L147 position is stabilized by a helical turn that immediately follows $\beta 5$ (Figure IV.5). The $\beta 6$ motif in most AdoMet radical enzymes involves peptide backbone positions at the end of $\beta 6$ to hydrogen bond with the adenine N1 and N6 positions. BtrN does not have a $\beta 6$ strand; instead, the backbone amide nitrogen and carbonyl groups of S150, the last residue in the protein's abridged AdoMet radical fold, hydrogen bonds to the adenine (2.9 and 3.1 Å from the N1 and N6 positions, respectively), much like traditional $\beta 6$ motif residues (Figure IV.6). All of these interactions make the AdoMet primary coordination sphere in BtrN very similar to those seen previously, thus one loop following $\beta 5$ in BtrN replaces two helices and a strand usually found in the AdoMet radical fold.

Following this modified 'β6 motif' in BtrN, R152 makes an interaction with AdoMet that is unique to previously characterized AdoMet radical enzymes (Figure IV.5). The residue is in a linker region (disordered in the OPEN structure) of the protein that joins the AdoMet radical domain and C-terminal domain. The guanidinium group of R152 stacks with the adenine base and interacts with the ring oxygen of the cofactor's ribose moiety. An arginine residue in the β6 motif of another AdoMet radical enzyme, spore photoproduct lyase, also contacts AdoMet, but it does not stack with the adenine or hydrogen bond to the ribose. Instead, it interacts with phosphate groups of substrate. R152 in BtrN also seems to play a role in substrate binding (discussed below). Following this site, a loop region (in which residues 161-164 cannot be resolved in the electron density) joins the AdoMet domain to the auxiliary cluster domain.

The C-terminal auxiliary cluster domain

The auxiliary cluster domain initiates with C169, the first auxiliary (Aux) cluster ligand. A β-hairpin follows before the second Aux cluster ligand position, C188, and an α-helical region (Figure IV.3). These secondary structural elements and ligand positions have very high structural similarity to the SPASM domain from anSMEcpe (17); residues C169 – E223 of BtrN align to residues C261 – F311 of anSMEcpe with an RMSD of 1.1 Å. The two proteins are very different after the α-helical region. BtrN terminates in a series of loops which contain the final two ligands to its Aux cluster (C232 and C235), while anSMEcpe continues, binding an additional [4Fe-4S] cluster (Figure IV.7). While this region does not align as well to MoaA (MoaA residues C264-I311; RMSD 4.4 Å) the iron ligating cysteine positions and secondary structural elements are conserved (Figure IV.7). Additionally, all three proteins share a conserved glycine residue in the β-hairpin motif between two auxiliary cluster ligands, in a $CX_{8-12}GX_4C$ motif (Figure IV.4). The effect of these instances of structural homology is that the AdoMet and Aux cluster positions in all three proteins are nearly identical; AdoMet cluster to Aux cluster distances in the three proteins are 15.9, 16.9, and 16.5 Å for BtrN, anSMEcpe, and MoaA, respectively.

DOIA binding

While we were unable to solve a crystal structure with DOIA bound using an C-terminally His_6 tagged BtrN construct, the substrate was readily apparent in the electron

density from N-terminally His$_6$ tagged BtrN crystals (Figure IV.8.C). DOIA binds in a hydrophilic pocket between the AdoMet and Aux clusters. The aminoglycoside is in a low energy chair conformation, with all five functional groups in equatorial positions. This conformation is stabilized by numerous hydrogen bonds to amino acid positions spanning the BtrN sequence, from E9 to V249 (Figures IV.8.A and IV.4). The final interaction is made between the amine position of DOIA and the amide nitrogen and carbonyl of the V249 peptide backbone, one residue from the C-terminus of the protein. The C-terminus, which itself provides a water-mediated hydrogen bond to DOIA (Figure IV.8.A), forms a loop that caps the underside of the BtrN active site. This capping function suggests that DOIA binds after AdoMet, consistent with previous substrate inhibition studies (5). While the C-terminally His$_6$ tagged BtrN is active (15), we believe that the tag impedes these capping interactions, thereby preventing the formation of a CLOSED conformation complex stable enough for structural characterization.

The C3 position of DOIA (the hydrogen abstraction site) is 3.7 Å away from the 5' carbon of AdoMet (Figure IV.8.A), consistent with prior structural characterizations of AdoMet radical enzymes (3.8–4.1 Å) (28). Following hydrogen abstraction, oxidation of the C3 and deprotonation of the C3-OH are required to complete turnover. Two titratable residues are present within 5 Å of C3-OH; H60 and R152 (Figure IV.8.A). H60 forms a hydrogen bond with the C4-OH, but does not interact with the C3-OH. R152 makes two 3.1 Å hydrogen bonds to the C3-OH position via both η nitrogens. R152 and a water (hydrogen bonding to another water and the F61 backbone carbonyl) are the only moieties making contact with this position of substrate. Possible electron transfer partners to facilitate oxidation of the C3 radical intermediate are both the AdoMet and Aux clusters. These clusters are 8.6 and 9.6 Å away, respectively, from the C3 position.

Sequence and structural homology to anSMEcpe

Although the active sites of BtrN and anSMEcpe show little sequence conservation (Figure IV.4), some residues important for substrate binding are found in similar positions in three dimensional space, with R152, a possible proton acceptor in BtrN, being the only major exception (see above). Based on mutagenesis studies, D277 was identified as a putative base in anSMEcpe (17). This residue is located on a loop in its auxiliary cluster domain next to F278, a residue that stacks against the Aux cluster (Figure IV.8.D). In BtrN, F188 aligns very well with D277 and stacks against its Aux cluster, which is in a slightly different position than in anSMEcpe. One residue downstream, E189 in BtrN overlays with F278 in anSMEcpe, and while

116

it does not interact with substrate, E189 forms a 2.7 Å hydrogen bond with R152. Thus, the residue may be indirectly involved in deprotonation of substrate. In anSMEcpe, two residues, Q64 and Q98 have different orientations in the presence and absence of substrates and are proposed play a role in substrate binding and orienting the putative base, D277 (17). In BtrN, the identities of these residues are H60 and Y90, which hydrogen bond to the C4 and C6 hydroxyl groups of substrate, respectively (Figure IV.8.D). These positions overlay very well in the OPEN and CLOSED BtrN structures, suggesting their orientation does not change upon binding substrate.

Y24 was another candidate for the general base in anSMEcpe based on its position in the structure. This residue resides in the AdoMet cluster binding loop proximal to the CX_2CX_3C motif. While mutational studies revealed that this residue most likely does not aide in the deprotonation of intermediate (17), we find a tyrosine residue in a similar position in BtrN. This tyrosine, Y62, is located on a loop following the β2 strand. Whereas Y24 in anSMEcpe interacts with the adenine base of AdoMet, in BtrN R152 wedges between AdoMet and Y62, separating it from both AdoMet and the substrate binding site (Y62 is 6.3 and 5.1 Å from AdoMet and DOIA, respectively, while in anSMEcpe, Y24 is 3.9 and 3.7 Å from AdoMet and substrate; Figure IV.8.D). E9 is another residue in BtrN that is much different in anSMEcpe (L7). The glutamate makes multiple hydrogen bonding interactions with the C4 and C5 hydroxyl groups of DOIA (both 2.7 Å) (Figure IV.8.D), while in anSMEcpe L7 provides a hydrophobic contact with a proline residue of the substrate peptide. All residues discussed in this section are marked by red stars in Figure IV.4. Interactions provided to DOIA by the C-terminal cap are also not conserved in anSMEcpe, where the sites of these interactions are substituted for the anSMEcpe peptide substrate itself.

IV.IV DISCUSSION

With structural information for BtrN (this work) and anSMEcpe (17), we have completed the structural characterization of the currently identified members of the AdoMet radical dehydrogenase subfamily. Using AdoMet radical chemistry, these enzymes catalyze nearly identical two electron oxidations (Figure IV.1) and exhibit different variants of the AdoMet radical fold. One, the partial TIM barrel fold found in anSMEcpe and common to nearly all AdoMet radical enzymes (27, 28), provides a highly conserved primary coordination sphere around AdoMet. In BtrN, however, this fold lacks 25% of the expected β_6/α_6 architecture secondary structure, adopting an even more abridged, β_5/α_4 partial barrel. This interesting

117

deviation contrasts with the striking level of structural similarity between the auxiliary cluster folds of BtrN and anSMEcpe. The two enzymes share an RMSD of 1.1 Å in this region even though BtrN binds one auxiliary (Aux) [4Fe-4S] cluster proximal to its active site, while anSMEcpe binds two, one near the active site and another near the protein surface (Figure IV.7). Despite this difference in FeS cluster content, the two proteins were hypothesized to use an active site auxiliary cluster for direct substrate ligation to catalyze similar reactions (9, 15, 18). We have previously indicated that this hypothesis is incorrect in the case of anSMEcpe (17), and we similarly show in this work that the BtrN Aux cluster is fully protein ligated in the presence and absence of substrate. Substrate is largely stabilized by interactions provided by three of the five C-terminal residues, potentially explaining the difficulty in crystallizing a substrate bound structure of a BtrN construct containing a C-terminal His$_6$ tag.

Direct ligation of the BtrN substrate, DOIA, to its Aux cluster was proposed to aide in deprotonating the DOIA C3 hydroxyl group and oxidizing the substrate radical intermediate (15). Without direct ligation to an auxiliary cluster to activate the hydroxyl group, deprotonation before AdoMet mediated hydrogen abstraction is unlikely due to the ~16 unit pKa of the C3-OH. Formation of the α-hydroxyalkyl radical by hydrogen abstraction at the C3 position will activate the hydroxyl functional group by decreasing its pKa by ~5 units (32). For subsequent deprotonation, we find only one titratable residue, R152, within 4 Å of the C3-OH group of DOIA. This residue makes contacts, 3.1 Å away, using both η nitrogens of the guanidinium group (Figure IV.8.A). While arginine sidechains have high pKas (~12.5) that are usually unsuitable for accepting protons during catalysis, nearby carboxyl side chains have been known to adjust arginine pKas for general base function (33). In BtrN, E189 forms a hydrogen bond / salt bridge with R152 (Figure IV.8.A), possibly activating it in this manner. Taken together, these results are consistent with R152 accepting a proton from the •C3-OH intermediate during catalysis.

After or concentered with deprotonation, collapse of the •C3-O⁻ radical to the C3=O product necessitates loss of an electron, most likely to one of BtrN's [4Fe-4S] clusters (5, 14, 15, 34). The AdoMet and Aux clusters are 8.6 and 9.6 Å, respectively, away from the C3 position of DOIA, making them both within range for a suitable electron transfer partner (35). A similar situation exists in anSMEcpe, where an oxidation event is required from a cysteinyl Cβ positioned nearly equidistant from its AdoMet and Aux clusters (8.9 and 8.6 Å, respectively) (17). Like in anSMEcpe, we suggest that the role of the Aux cluster in BtrN is to accept an electron from substrate during turnover. In anSMEcpe, the reduced active site Aux cluster is positioned to reduce a second Aux cluster at the protein surface for transfer to a terminal

electron acceptor. This pathway appears to be required in anSMEcpe because the large peptidyl substrate in the reaction buries the anSMEcpe active site Aux cluster, necessitating transfer to a more solvent exposed cluster before transfer to a partner protein (17). The small molecule substrate of BtrN, in contrast, does not change the solvent accessibility of the Aux cluster, removing the need for a second Aux cluster. To reoxidize the Aux cluster in BtrN, electron transfer to the AdoMet cluster has been proposed, which would ready this cluster for the next round of catalysis (15). At 15.9 Å, however, the two clusters are too far apart for unmediated electron transfer, suggesting that, like in anSMEcpe, electron transfer between the Aux and AdoMet clusters does not occur (17). We therefore propose that following turnover in BtrN, a partner protein reoxidizes the reduced Aux cluster directly, without an internal electron transfer event.

Another carbohydrate modifying AdoMet radical protein, DesII, also has interesting redox requirements. In considering it's native deaminase activity (Figure IV.1), the removal of an amine group from TDP-4-amino-4,6-dideoxy-D-glucose is coupled to the oxidation of an alcohol to a ketone. While this overall reaction is net redox neutral, it requires the *reduction* of a substrate radical intermediate following hydrogen abstraction (36). Interestingly, DesII can also catalyze a dehydrogenase reaction when given the unnatural substrate, TDP-D-quinovose. This reaction, like BtrN, requires the *oxidation* of the substrate radical intermediate, which was found by kinetic isotope effects to occur either following or concerted with deprotonation (34) (Figure IV.1). In both cases, the AdoMet cluster is hypothesized to be the electron transfer partner as DesII does not have any auxiliary clusters (37, 38). How DesII accomplishes the same chemistry as BtrN (and anSMEcpe) in the absence of an auxiliary cluster could be explained by the DesII AdoMet cluster's innate ability to react with substrate. Supplying an unnatural substrate with a poor leaving group presumably prolongs the lifetime of the potent •C3-O⁻ species ($E°' < -1.6$ V), reversing the flow of electrons relative to the native reaction (32). In the BtrN and anSMEcpe reactions, nature has made available an auxiliary cluster, which, in BtrN cannot be reduced by titanium citrate ($E°' \sim -480$ mV), implying the cluster has a low redox potential that is tuned to match that of the substrate radical intermediate (15). These hypotheses await further electrochemical characterization of the clusters.

Two mechanisms have been hypothesized for the DesII elimination of the amine moiety in the native reaction. One proposal is that the C4-NH_3 group is eliminated in a E1cb-type mechanism, creating a enol radical species that, upon collapse of the ketone, can protonate and reduce to the product (39). This mechanism relies on the ability of the enzyme to active the leaving group, forming a C4-NH_3^+ species. The second hypothesis, reminiscent of B_{12} dependent

radical mechanisms, suggests that the amine group migrates to a •C3 intermediate, forming a carbanolamine radical species that readily deaminates forming a ketone (39). This mechanism also requires protonation/reduction and relies on a high degree of overlap between the radical and C4-NH$_3$ orbitals for efficient migration (40). When given the unnatural TDP-D-quinovose substrate, EPR studies have indicated little orbital overlap between these two positions (37). It is unknown, however, what configuration substrate assumes in the native DesII reaction and the environment surrounding the C4 functional group.

In contrast, the structure of BtrN shows that the enzyme is well designed to avoid an elimination reaction by either mechanism. First, an E1cb-type mechanism is disfavored, as the C4-OH group is not activated for elimination. The functional group is involved in two tight (2.7 Å) hydrogen bonds with the E9 sidechain, which itself is being activated as a hydrogen bond acceptor through interactions with a nearby arginine (R56, Figure IV.8.A). Second, elimination via a migration-assisted mechanism is not feasible as DOIA binds in an equatorial chair conformation with little overlap between the hydrogen abstraction site and the C4-OH bonding orbital (Figure IV.9.A). The substrate is held in this conformation by multiple hydrogen bonds and van der Waals interactions. Modeling a partial boat DOIA conformation to increase orbital overlap distorts the active site hydrogen bonding network and introduces steric clashes with Y90 and F188, two residues positioned above and below the DOIA cyclohexane ring (Figure IV.9.B). Thus, BtrN disfavors a DesII-like elimination, given either mechanism. The existence of the Aux cluster in BtrN, which we suggest drives oxidation of the radical intermediate, further distinguishes it from DesII.

The Aux cluster binding domains of BtrN, anSMEcpe, and MoaA place them in a large subfamily within the AdoMet radical superfamily. AnSMEcpe is a member of the recently described SPASM subfamily (17, 18), whose members contain two [4Fe-4S] clusters in a C-terminal 7-cysteine motif. BtrN only binds one cluster using a cysteine motif similar to the first half of the SPASM motif, which we have termed a twitch domain. We now conclude that this twitch domain comprises the β1′/ β2′ hairpin and the α2′ helix (Figure IV.10). In addition to BtrN, MoaA (23) contains a twitch domain with the same C-terminal secondary structural elements as BtrN and anSMEcpe (Figures IV.10 and IV.7). Mapping these three proteins onto a representative diagram of the AdoMet radical superfamily (Figures IV.11 and IV.12), suggests that this twitch/SPASM domain will be a common feature to ~ 30% of uncharacterized AdoMet radical enzymes. While it is difficult to predict Aux FeS cluster content in these family members, all currently characterized members of this twitch/SPASM cluster contain at least one Aux cluster (8, 9, 15, 17, 19-21, 41).

Contained in this subfamily is the recently described thioether formation class of AdoMet radical enzymes, represented by AlbA (19) and SkfB (41) (Figure IV.11). SkfB has a C-terminal motif with five cysteines, which we suggest forms a twitch domain much like that of BtrN (Figure IV.10). AlbA, a founding member of the SPASM subfamily, contains seven cysteines in its C-terminal domain, in an arrangement much like anSMEcpe (Figure IV.10). Upon removal of the AdoMet cluster, AlbA still has [4Fe-4S] cluster EPR signal, which was attributed to one auxiliary cluster (19). Based on sequence analysis, we suggest, that AlbA binds two auxiliary [4Fe-4S] clusters in a SPASM arrangement similar to that of anSMEs. Like BtrN and anSMEs, AlbA and SkfB catalyze very similar reactions, namely, the linkage of a peptidyl cysteine thiol to the Cα position of an up or downstream residue on the same peptide substrate. We propose that this system will be another example of a SPASM containing protein (AlbA) and a twitch domain containing protein (SkfB) catalyzing similar chemistry on different substrates.

The positioning of enzymes in the SAPSM/twitch sequence cluster is consistent with our auxiliary cluster architecture designations. The SPASM proteins (anSMEs and green nodes in Figure IV.11) are localized on one end of the cluster, while the twitch proteins, BtrN and MoaA, are located in smaller subclusters. AlbA and SkfB, however, map nearly on top of each other, right where we believe the junction exists between SPASM and twitch containing nodes. Functionally, whether the direct ligation feature of MoaA is conserved in SkfB, AlbA, or any other SPASM/twitch subclass member is an interesting issue and awaits further biochemical and structural studies. Regardless, the structures of BtrN provide the third example of a highly conserved Aux cluster binding architecture, suggesting that it will be present in the thioether bond forming enzymes and across this abundant and interesting SPASM/twitch AdoMet radical subclass.

IV.V   MATERIALS AND METHODS

Protein purification and production of DOIA were performed as described (15).

Crystallization

To crystallize C-terminally His$_6$ tagged SeMet derivitized BtrN, a protein solution (containing 48 mg/mL C-terminally His$_6$ tagged BtrN, 10 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, and 5 mM DTT) was diluted in water to a final concentration of 20 mg/mL protein. AdoMet, generated enzymatically as previously described (42), was added to the solution to a

final concentration of 1 mM. This mixture was allowed to incubate for 16-24 hours at room temperature in an anaerobic environment (95% Ar, 5% H_2, COY Laboratory Products, Inc.). Following incubation, crystal drops using the sitting-drop diffusion technique were set up by combining 1 μL of the above protein solution with 1μL of a precipitant solution containing 13-18 % Jeffamine ED-2001 pH 7.0 (Hampton Research) and 100 mM imidazole pH 7.0. Rod-shaped crystals would grow in 3-7 days with dimensions of approximately 20 x 20 x 100 μm. Crystals were looped and transferred into a drop of Paratone-N (Hampton Research). This cryoprotectant was effective at shrinking the unit cell by dehydrating the crystal, leading to a more ordered lattice. After vigorous rinsing to fully exchange the mother liquor, the crystals were looped and cryocooled by direct submersion into liquid nitrogen.

To crystallize N-terminally His_6 tagged native BtrN, a protein solution (containing 138 mg/mL N-terminally His_6 tagged BtrN, 10 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, and 5 mM DTT) was diluted in water to a final concentration of either 15 or 20 mg/mL. AdoMet and DIOA (from where?) were added to a final concentration of 5 mM. This solution was incubated at room temperature for 4-8 hours and then used to set up hang-drop diffusion method drops with 1 μL of the above protein solution and 1μL of a precipitant solution containing 30 mM citric acid / 70 mM Bis-TRIS propane pH 7.6 and 25-31% PEG 4000 (Hampton Research). Crystals would appear overnight and grow to dimensions of 50 x 50 x 300 μm in 3-7 days. For cryocooling, crystals were looped and washed through a drop of precipitant solution containing 5 mM AdoMet, 5 mM DOIA, and 20% glycerol, then submerged in liquid nitrogen.

Both the N-terminal and C-terminal His_6 tags could not be cleaved due to the lack of a protease site in the constructs. Crystallization of native enzyme (without an affinity tag) was attempted, but unsuccessful presumably due to heterogeneity.

Data Collection and Structure Determination

An initial dataset of C-terminally His_6 tagged SeMet BtrN (Se-SAD, Table IV.1) was collected at 100 K using inverse beam (1° oscillations, 30° wedges) at the selenium peak (0.9792 Å) to 2.50 Å resolution. As anomalous signal was present to better than 3 Å (phenix.xtriage), the AutoSol Wizard (43) was used to the full resolution of the dataset. Eleven heavy atom sites were found with a 0.48 figure of merit to 2.50 Å resolution. With one molecule in the asymmetric unit, these sites correspond to the 10 methionine positions in BtrN and to a peak from the BtrN auxiliary [4Fe-4S] cluster. Experimental maps were solvent flattened using RESOLVE, yielding

interpretable electron density. Both [4Fe-4S] clusters and all residues (including sidechains) except for 1-3, 25-35, 116-166, and 234-250 were built into the experimental density. Using isomorhous replacement, this model was used to phase a higher resolution dataset (to 2.02 Å) collected on another SeMet BtrN crystal (Hi-res SeMet, Table IV.1). Using these data, the model was completed, however electron density for residues 28-32, 121-134, 146-161, 250 and the 19-residue C-terminal His$_6$ tag remained negligble. This structure was refined using 6 TLS groups and no sigma cutoff.

A dataset of N-terminally His$_6$ tagged native BtrN (AdoMet/DOIA, Table IV.1) was collected at 100 K using sequential 0.5° oscillations to 1.56 Å resolution. These data were phased using molecular replacement with the Hi-res SeMet model (Phaser Z-score of 21.7 using a 4 Å resolution cutoff) (43). Electron density for missing regions of the Hi-res SeMet model is apparent using these data, including both AdoMet and DOIA substrates and the full protein chain (except for residues 161-164). The N-terminal His$_6$ tag was completely disordered. Data for all structures were collected on either beamline 24-IDE or 24-IDC at the Advanced Photon Source (Argonne, Il) and processed in HKL2000 (44). All models were refined and built in PHENIX (45) and COOT (46), respectively. Composite omit maps were used to verify the final models. See Table IV.1 for full data processing, refinement, and validation statistics.

IV.VI ACKNOWLEDGEMENTS

Author contributions

P.J.G., T.L.G, S.J.B., and C.L.D. designed research; P.J.G. performed the crystallization and crystal structure determination; T.L.G. provided protein samples for crystallization; and P.J.G. and C.L.D. wrote the paper.

Table IV.1. Data processing and refinement statistics for BtrN.

	Se-SAD SeMet (C-term His$_6$)[1]	Hi-res SeMet (C-term His$_6$)[1]	AdoMet/DOIA (N-term His$_6$)
Beamline	24-IDE	24-IDC	24-IDC
Data Processing			
Wavelength (Å)	0.9792	0.9795	0.9795
Space Group	C2	C2	$P2_12_12_1$
Cell Dimensions			
a, b, c (Å)	99.6, 42.1, 50.9	99.4, 42.1, 50.9	44.6, 55.2, 116.9
β (°)	104.9	105.4	90.0
Resolution (Å)[2]	50.0 – 2.50 (2.59 – 2.50)	50 – 2.02 (2.09 – 2.02)	100 – 1.56 (1.62 – 1.56)
R_{sym} (%)[2]	12.4 (37.9)	8.7 (42.9)	11.6 (46.4)
$<I/\sigma(I)>$[2]	21.5 (4.8)	20.6 (3.0)	10.7 (2.0)
Completeness (%)[2]	99.9 (99.1)	97.1 (74.0)	95.7 (76.0)
Redundancy[2]	7.6 (5.3)	5.6 (4.2)	3.9 (2.6)
Total reflections	105,617	143,572	158,698
Refinement			
Resolution (Å)		38.5 – 2.02	58.4 – 1.56
Reflections		13,349	41,942
R_{work} / R_{free}[3]		16.5 / 21.6	16.3 / 18.6
No. of non-hydrogen atoms			
Protein		1,699	1,949
[4Fe4S] / AdoMet / DOIA		16 / - / -	16 / 27 / 11
Solvent		74	281
Average B-factors (Å^2)			
Protein		37.4	19.8
[4Fe4S] / AdoMet / DOIA		28.2 / - / -	15.9 / 16.4 / 14.5
Solvent		43.8	22.5
RMS deviations			
Bond lengths (Å)		0.015	0.007
Bond angles (°)		1.51	1.19
Ramachandran statistics[4]		86.5%, 13.5%, 0%, 0 residues	90.0%, 9.6%, 0%, 1 residue

[1]Data were scaled anomalously
[2]Highest resolution shell is shown in parentheses
[3]R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma|F_{obs}|$ and R-free is the R value for a test set of reflections consisting of 5% of the diffraction data not used in refinement.
[4]Values reported correspond to the number of residues in the most favored, additionally allowed, generously allowed, and disallowed regions, respectively.

Figure IV.1. Butirosin B biosynthesis and AdoMet radical dehydrogenase activity in BtrN, anSMEs, and DesII; and the native DesII deaminase activity.
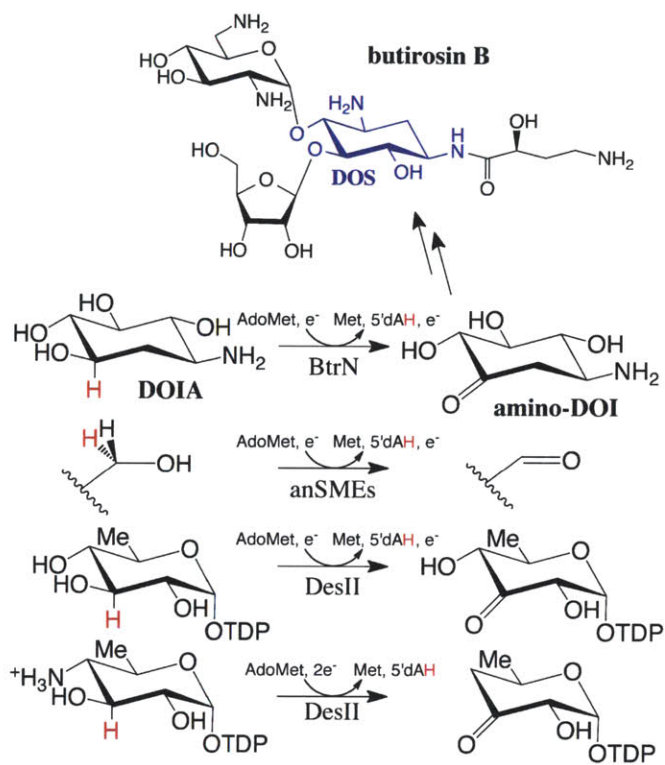
Figure IV.2. OPEN and CLOSED conformations of BtrN. OPEN (tan) and CLOSED (grey) conformations of BtrN (RMSD 0.4 Å). AdoMet and DOIA (shown in sticks) are bound to the CLOSED conformation. Both models contain two [4Fe-4S] clusters. *(A)* Slight changes in the positioning of the AdoMet cluster binding loop are due to the large difference in orientation of W21. Additionally, a linker region which bridges the AdoMet and Aux cluster domains (see text for domain definitions) and provides contacts to the adenine of AdoMet, is disordered in the OPEN structure (after the linker region, three residues are disordered in the CLOSED structure). *(B)* While the β4 and β5 strands are superimposable in the two structures, α4 is disordered in the OPEN structure and not included in the model. *(C)* The C-terminal cap in the OPEN structure leaves the DOIA binding site exposed, while in the CLOSED conformation, this region makes many contacts to the substrate (see Figure IV.8). The arrow indicates the position of the V249 Cα in the OPEN and CLOSED structures (residue 250 is only modeled in the CLOSED structure).
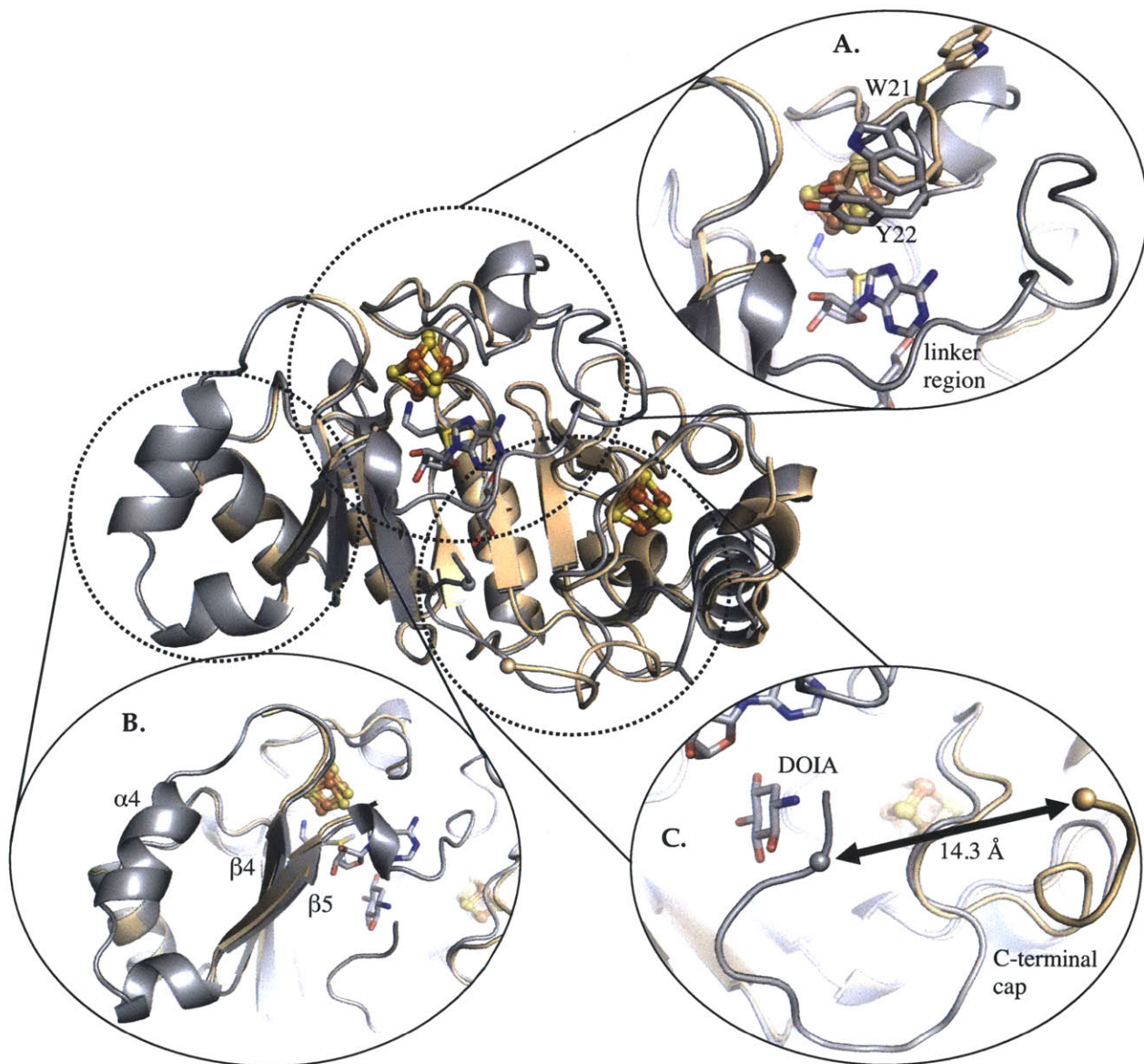
Figure IV.3. The BtrN fold. (*A*) BtrN includes an AdoMet domain (residues 1-150; magenta), an auxiliary cluster domain (residues 169-235; green), and linker regions (residues 151-168 and 236-250; light blue). (*B*) The BtrN substrate, DOIA (grey), binds between the two clusters. (*C*) Topology of BtrN.
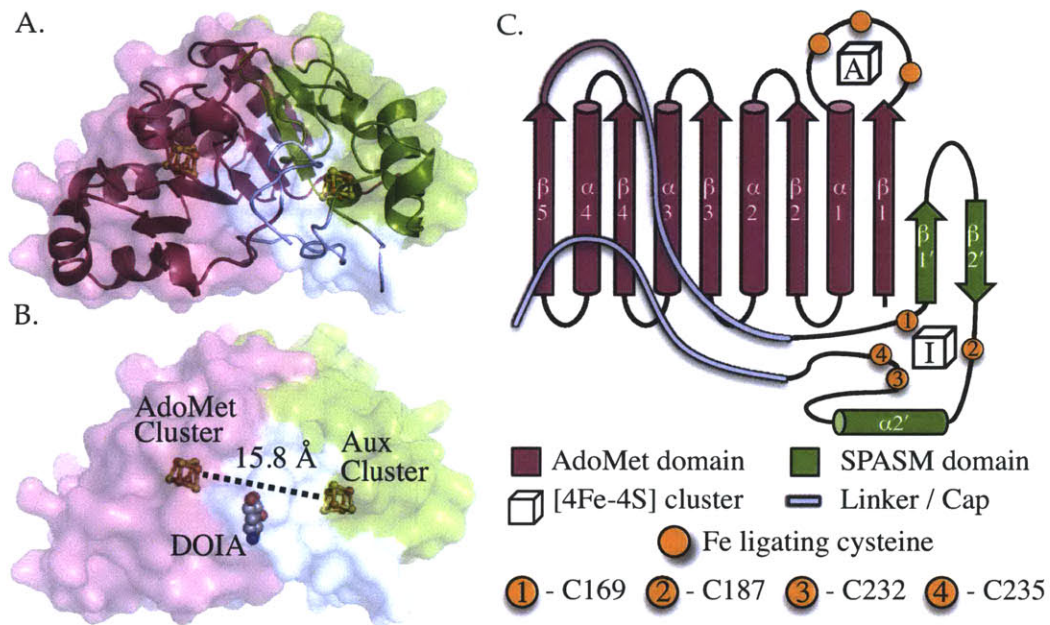
Figure IV.4. Structure based sequence alignment of BtrN and anSMEcpe. Secondary structure assignments are colored green for β strands and yellow for α helices. The AdoMet fold (residue 1-150 in BtrN and 1-234 in anSMEcpe) and the 'twitch' domains, including β1', β2', and α2' (residues 169-235 in BtrN and 255-311 in anSMEcpe) align well. The RMSD histogram (blue) denotes Cα distances between aligned residues and ranges from 0.05 Å (thinnest bar) to 3.9 Å (thickest bar); overall RMSD of 185 aligned residues is 2.02 Å. Strands, helices, and AdoMet radical motifs are labeled beneath the alignment. Residues that contact AdoMet are in bold and red stars denote substrate binding residues discussed in the Results section. Black arrows indicate Aux cluster Fe ligating cysteines (grey arrows indicate cysteines ligating anSMEcpe's second auxiliary cluster). Alignment generated by Chimera (47).

```
RMSD
Consensus   . . . . . MDKLFS   sieikpaS.q   CNrkCtYCp.   h.s...rskk   .syGiMrpal
BtrN      1 . . . . MDKLFS      MIEVEVNS.Q   CNRTCWYCP.   N.S...VSKR   KETGEMDPAL
anSMEcpe  1 MPPL. . . . . .     SLLIKPASSG   CNLKCTYCFY   HSLSDNRNVK   .SYGIMRDEV
                               β1 ☆        AdoMet radical motif ☆

RMSD
Consensus   yksmmkqlss   ea...acsia   Fqg.GEPtla   k..nlekl.m   mg.r..y..
BtrN     41 YKTLMEQLSS   LD..FAGRIS   FHFYGEPLLC   K..NLDLFVG   .MTTE..Y.
anSMEcpe 44 LESMVKRVLN   EANG.HCSFA   FQG.GEPTLA   GLEFFEKL.M   ELQ.RKHNYK
            α1           β2        ☆ ☆'GGE' motif      α2

RMSD
Consensus   ....kiypii   qTNGtlid.k   .raqtLsE.g   .i.kfilsq.
BtrN     82 ..IPRARPII   YTNGDFLTEK   RLQTLTELG    .IQKFIVTQ.   .........H
anSMEcpe 90 NL..KIYNSL   QTNGTLID.E   SWAKFLSE.N   KF.LVGLSMD   GPKEIHNLN.
            β3 ☆        α3           β4 ribose motif  α4a

RMSD
Consensus   ak.....a.k   .h.fr.r..y   .......k...   k...kvv.i.   ...cl...
BtrN    118 AG.....A.K   .HKFRGV..Y   ...DQL.AGA    D.KEKVVYL.   DHSDL....
anSMEcpe 134 RKDCCGLDTF  SK.VE.RAAE   LFK..KY...    KV..EFN.IL   ...CVVTSNT
            α4                       β5         'GXIXGXXE' motif

RMSD
Consensus   ..........   ..........   .vcsd...l    .........y   ..i.p.....
BtrN    148 ..........   ....VLSN.RGG              ......ILDN   ..I.P.....
anSMEcpe 171 ARHVNKVYKY  FKEKDFKFLQ   FINCLDP..L   YEEKGK...Y   NYSLKPKDYT
            α5                      β6  β6 motif

RMSD
Consensus
BtrN    161 ..........                                          ...QASKANM
anSMEcpe 216 KFLKNLFDFW  YEDFLNGNRV   SIRYFDGLLE   TILLGKSSSC   GMN......
            α6

RMSD
Consensus   s..C..tc.q   aVVesdGsVy   PCdey...q.   KmrmGNIqeq   hikeiwhtdK
BtrN    168 S..CMVPSNL   AVVTVLGNVL   PCFED.FNQ.   KMVMGNIGEQ   HISDIWHNDK
anSMEcpe 259 GT.C..TC.Q   FVVESDGSVY   PCDFYV..LD   KWRLGNIQDM   TMKELFETNK
            ʌ            β1'     β2'  ʌ ☆☆

RMSD
Consensus   fhsFiKms.e   ..........   ..........   ..........   ..ks.....
BtrN    214 FTSFRKMLKE   ..........   ..........   ...GHRGKSD   LCKN......
anSMEcpe 303 NHEFIKLS.F   KVHEECKKCK   WFRLCKGGCR   RCR.......   ..DSKEDSAL
            α2'          ʌ   ʌ        ʌ    ʌ                   ʌ

RMSD
Consensus
BtrN    235 ..........   ........CN   NVSVQTEEQY   DYVL
anSMEcpe 343 ELNYYCQSYK   EFFEYAFPRL   INVANNIK..   ..........
            ʌ                         ʌ
```

Figure IV.5. AdoMet binding motifs in BtrN. Shown in stereo, BtrN (colored as in Figure IV.3, rotated 90° counterclockwise from Figure IV.8) binds AdoMet (grey sticks) using conserved AdoMet radical structural elements, including: coordination of the amine and carboxyl groups of AdoMet by the unique iron in the AdoMet [4Fe-4S] cluster (ligated by C16, C20, and C23); hydrogen bonding of the backbone of an aromatic residue in the CX_3CX_2C motif to the adenine (Y22 in BtrN); the 'GGE' motif (Y62, G63, and E64), in which backbone oxygen atoms hydrogen bond with the methionyl amine group of AdoMet; and the ribose motif (T115 and H117), in which polar residues at the end of the β4 strand hydrogen bond with the ribose hydroxyl groups. In some cases, β4 residues, such as H117, will also hydrogen bond to the carboxyl group of AdoMet (see text). The 'GXIXGXXE' motif (L147, stabilized by a turn that includes H114), which provides hydrophobic contacts to the adenine ring, and the 'β6 motif' (S150), which provides backbone hydrogen bonds to the adenine moiety are abridged in BtrN. (see Figure IV.8) (27, 28).
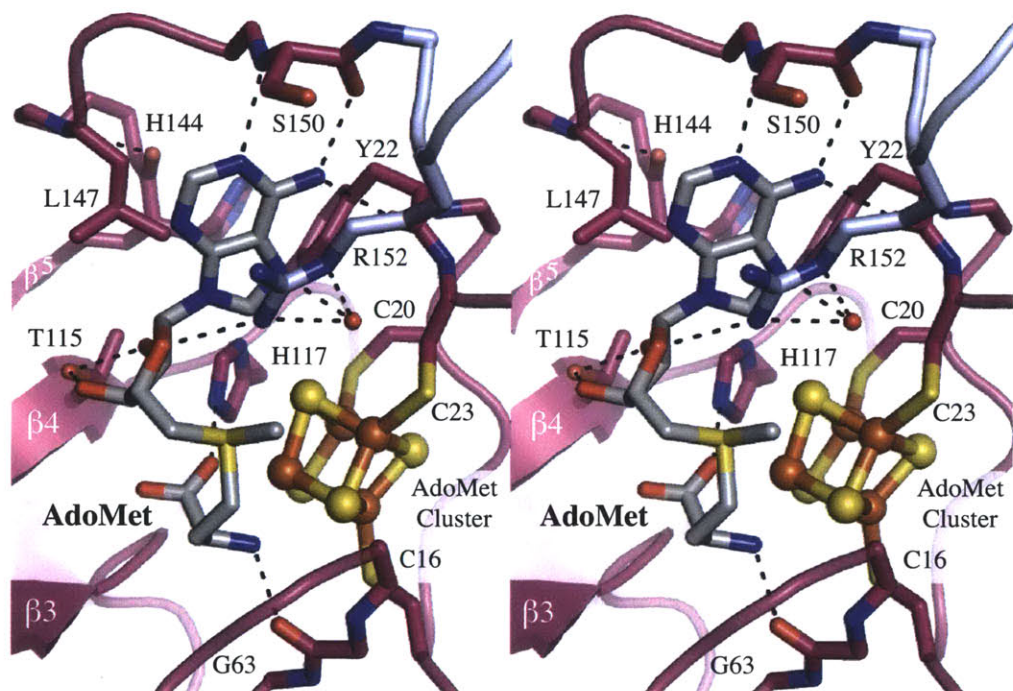
Figure IV.6. The abridged BtrN AdoMet radical fold. BtrN (colored as in Figure IV.3) lacks the α5 and α6 helices and the β6 strand found in all structurally characterized AdoMet radical proteins (represented by anSMEcpe in tan) (17). The 'GXIXGXXE' motif provides a hydrophobic contact to the AdoMet adenine and stabilization to the loop after β5 (V165, T167 and T170 in anSMEcpe). This hydrophobic contact is conserved in BtrN (L147) and stabilized by an upstream residue (H144, see Figure IV.5). Contacts provided by the backbone atoms of residues in the β6 motif of AdoMet radical enzymes (L195 in anSMEcpe) are conserved in BtrN (S150; Cα's in spheres), but are found in a loop following the β5 strand. R152 (Cα in light blue sphere) is unique to BtrN (see Figures IV.5 and IV.8).
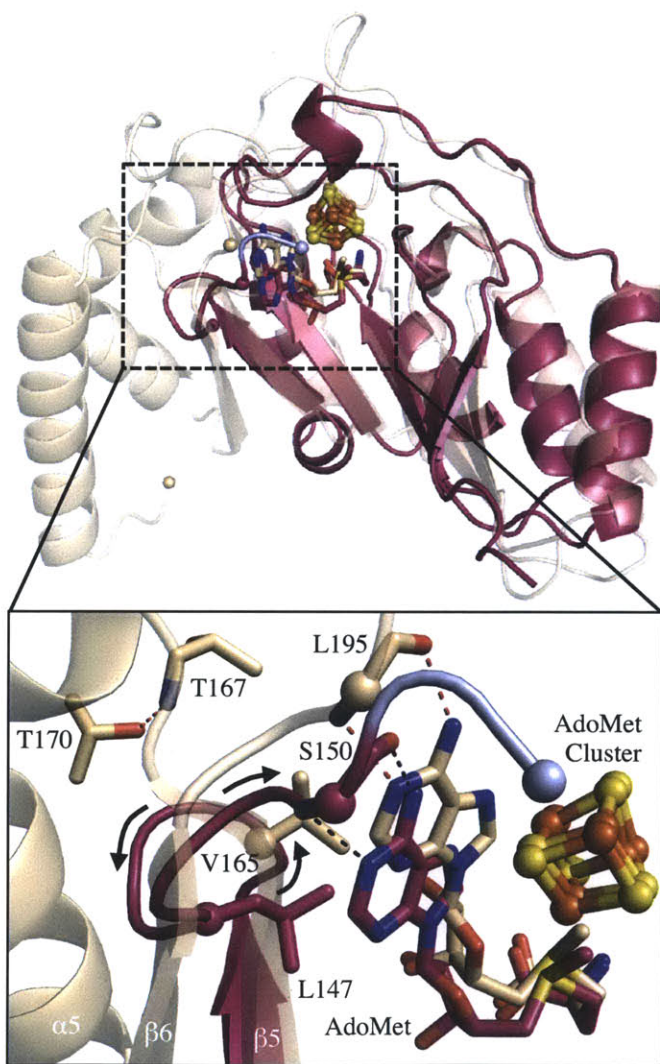
Figure IV.7. Variation of the AdoMet and Auxiliary cluster fold in BtrN, compared to anSMEcpe and MoaA. Auxiliary cluster domains in BtrN (green) compared to (*A*) anSMEcpe (tan), and (*B*) MoaA (grey). Black spheres denote iron ligating cysteine positions. The β1', β2', and α2' secondary structure elements are conserved in the three proteins (see Figure IV.10).
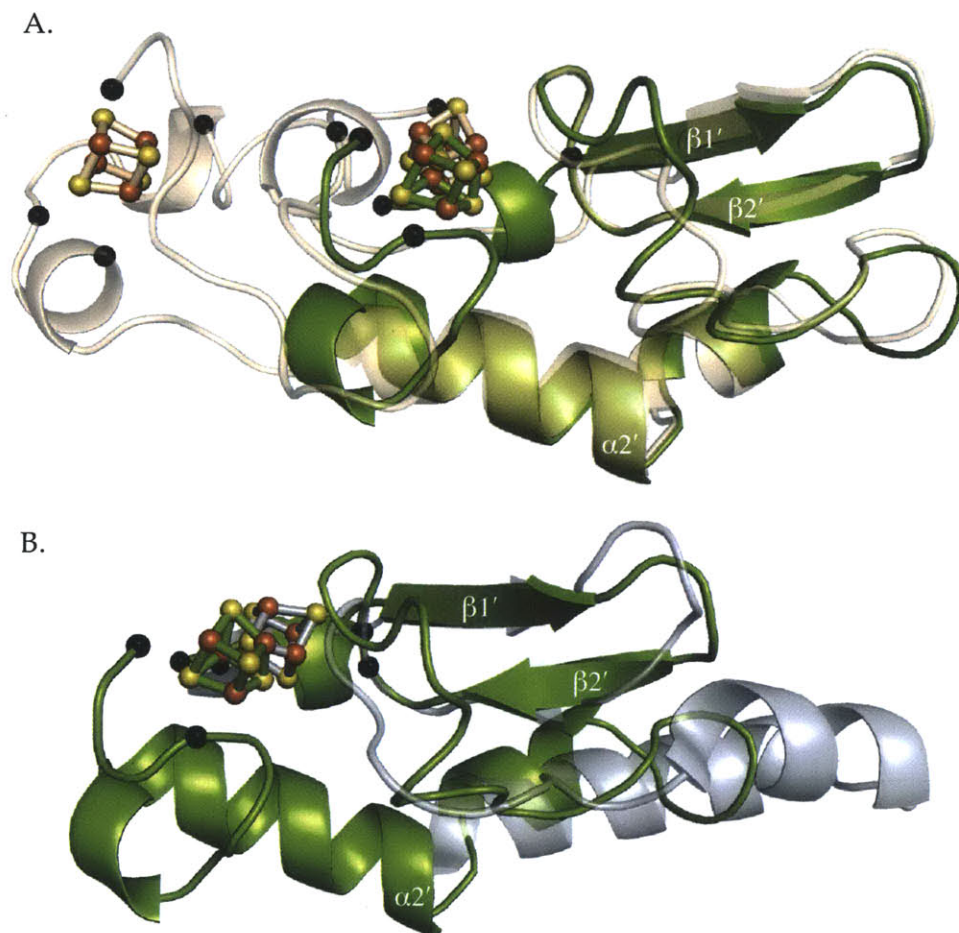
Figure IV.8. Substrate binding in the AdoMet radical dehydrogenases. (*A*) Residues contacting DOIA (grey) in BtrN (colored as in Figure IV.3). Red dashes indicate the separation of the 5′ position of AdoMet and the DOIA hydrogen abstraction site, C3. (*B*) $2F_o\text{-}F_c$ composite omit electron density countered at 1.0 σ for AdoMet and the AdoMet cluster and (*C*) substrate. (*D*) An overlay of the BtrN (colored as in *A*) and anSMEcpe (tan) active sites. AnSMEcpe sidechains are in parentheses. The 5′ position of AdoMet and hydrogen abstraction positions of the two substrates, DOIA and substrate peptide, are in spheres.
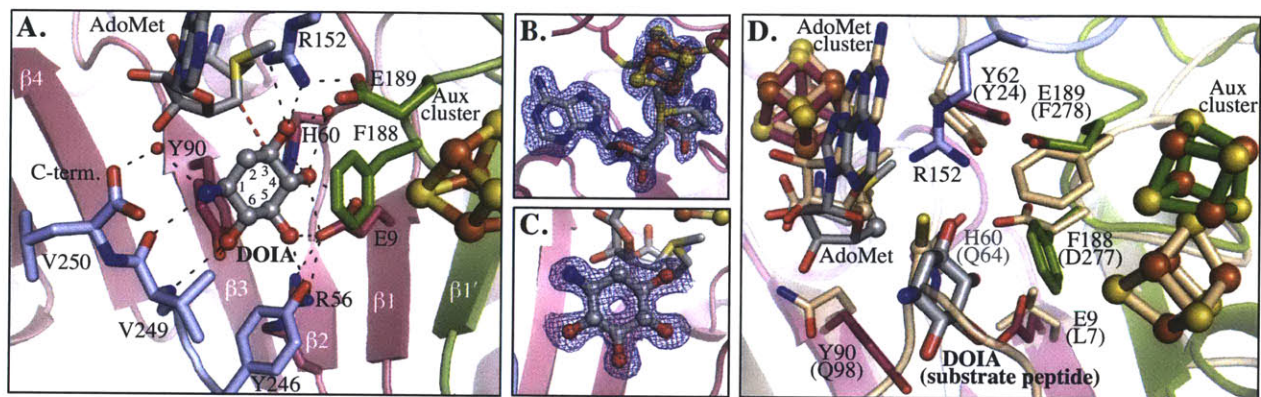
Figure IV.9. Possibility of alternative DOIA conformations in BtrN. (*A*) DOIA as found in the BtrN active site (colored as in Figure IV.8, C3, C4, and C5 positions labeled). Hydrogen bonds between the E9 and the C3-OH and C4-OH of BtrN (both 2.7 Å) are shown in black dashes. The C3-hydrogen is modeled in sticks (white) and the distance between it and the 5′ carbon of AdoMet is indicated by red dashes. (*B*) DOIA with a partial boat pucker is modeled into the DOIA active site. The C3 positions of both DOIA molecules are equivalent. Hydrogen bonding to E9 is no longer observed and clashing interactions between F188 and the C5-OH, and between Y90 and the C4-OH positions are now present.
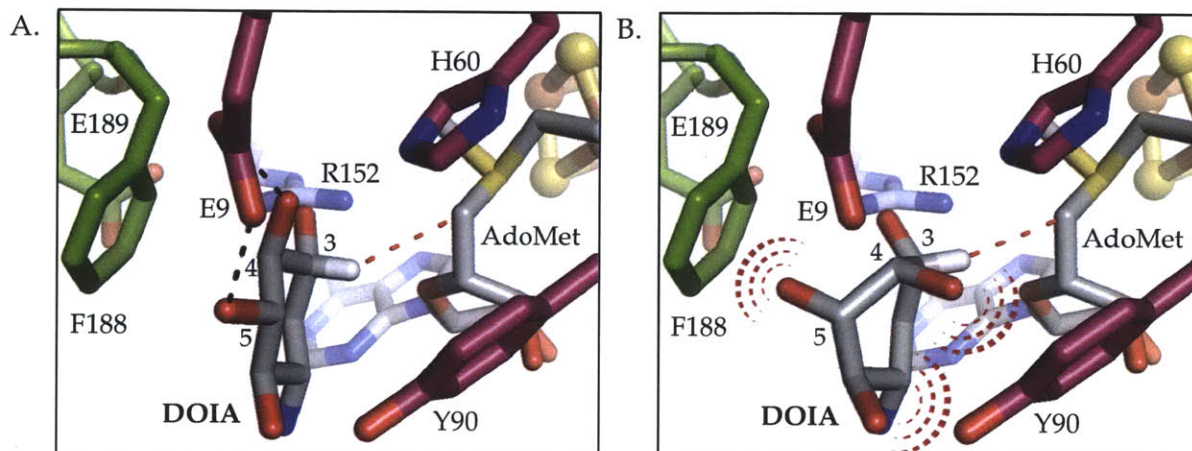
Figure IV.10. Auxiliary cluster sequence motifs. (*A*) Topologies of the Auxiliary cluster domains in BtrN, MoaA, and anSMEcpe (iron ligating cysteines in orange; MoaA substrate in yellow). (*B*) C-terminal domains in five auxiliary FeS cluster containing AdoMet radical proteins. Orange circles represent characterized (BtrN, MoaA, anSMEcpe) or possible (Alba, SkfB) iron ligating cysteine positions. In green are structural elements found in BtrN, MoaA, and anSMEcpe. In red are conserved elements of the SPASM and twitch domains (17).

A.

B.

	$\beta 1'$	$\beta 2'$	$\alpha 2'$	
BtrN	CMVP---SNLAVVTVLGNVLPC----X_{35}----EGHRGKSDL--CKN-C--X_{15}-			
MoaA	CS---TCTRARL-SSDGKFYGC----X_{62}----			
anSMEcpe	CGMNGTCTCQFVVESDGSVYPC----X_{40}----C--KKCKWFRLCKGGC--X_{17}-C-X_{22}-			
AlbA	CGAG---WKSIVISPFGEVRPC----X_{40}----CMKDKCPFSGYC-GGC--X_{15}-C-X_{15}-			
SkfB	CTPG---YLAWYIRADGYVTPC----X_{28}--------CEAKN-CK--C--X_{23}-			
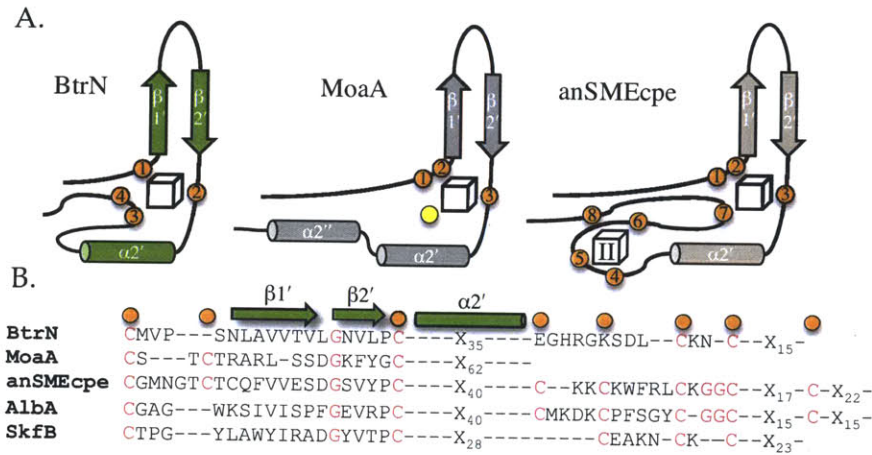
Figure IV.11. The AdoMet radical twitch/SPASM cluster. (*A*) The SPASM/twitch cluster with twitch members (BtrN, MoaA, SkfB) and SPASM members (anSMEs, AlbA) labeled. (*B*) The entire superfamily (nodes are enzymes with 40% sequence identity) using a threshold cut-off of an *E*-Value of 1×10^{-22}. Figure generated using Cytoscape (48). A more detailed figure of the superfamily can be found in Figure IV.12.
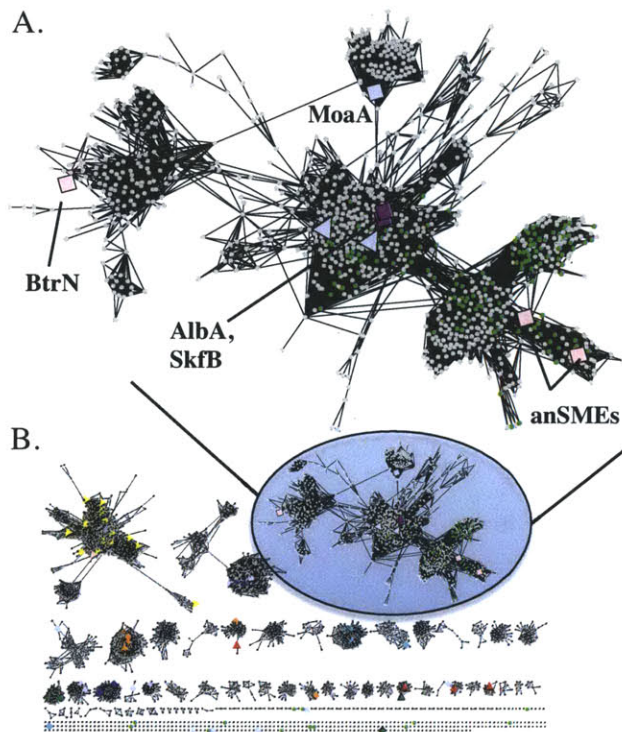
Figure IV.12. The AdoMet radical superfamily. A modification of the Enzyme Function Initiative's Radical SAM Superfamily Workshop materials (http://enzymefunction.org/resources/workshops). The entire superfamily (nodes are enzymes with 40% sequence identity) is shown using a threshold cut-off of an *E*-Value of 1×10^{-22}. Figure generated using Cytoscape (48).

○ Dehydrogenases
● Thioether Bond Formation
○ Complex Rearrangement
● SPASM Domain Proteins (TIGR04085); Reactions Unknown
● B12-like Rearrangement
○ Glycyl Radical Activating enzymes
● Metallocofactor Assembly
● Methylation Reactions (B12 Dependent)
● Methylation Reactions (non-B12)
● Sulfur Insertion
○ Reactions Unknown

○ Structure Available
△ Experimental Evidence
◇ Predicted

Notable Protein Nodes

1.	BtrN	○	(twitch)
2.	anSMEs	○	(SPASM)
3.	MoaA		(twitch)
4.	DesII		
5.	AlbA	●	(SPASM)
6.	SkfB	●	(twitch)
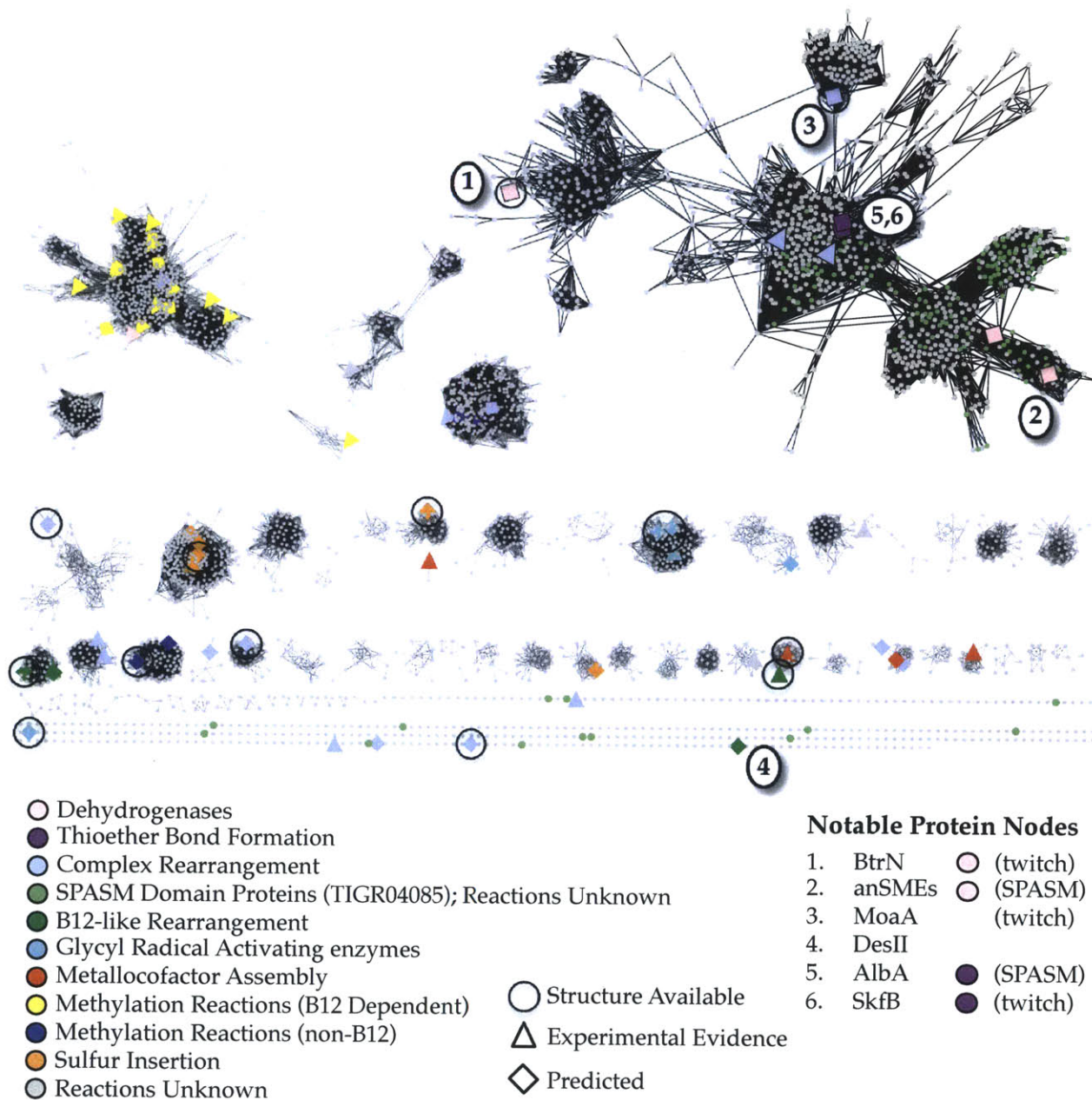
IV.VII REFERENCES

1. Davies J & Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74(3):417-433.

2. Vakulenko SB & Mobashery S (2003) Versatility of aminoglycosides and prospects for their future. *Clin Microbiol Rev* 16(3):430-450.

3. Wehmeier UF & Piepersberg W (2009) Enzymology of aminoglycoside biosynthesis-deduction from gene clusters. *Methods Enzymol* 459:459-491.

4. Kudo F, Yamamoto Y, Yokoyama K, Eguchi T, & Kakinuma K (2005) Biosynthesis of 2-deoxystreptamine by three crucial enzymes in Streptomyces fradiae NBRC 12773. *J Antibiot (Tokyo)* 58(12):766-774.

5. Yokoyama K, Numakura M, Kudo F, Ohmori D, & Eguchi T (2007) Characterization and mechanistic study of a radical SAM dehydrogenase in the biosynthesis of butirosin. *J Am Chem Soc* 129(49):15147-15155.

6. Kudo F & Eguchi T (2009) Biosynthetic genes for aminoglycoside antibiotics. *J Antibiot (Tokyo)* 62(9):471-481.

7. Frey PA, Hegeman AD, & Ruzicka FJ (2008) The Radical SAM Superfamily. *Crit Rev Biochem Mol Biol* 43(1):63-88.

8. Benjdia A, et al. (2008) Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. *J Biol Chem* 283(26):17815-17826.

9. Grove TL, Lee KH, St Clair J, Krebs C, & Booker SJ (2008) In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry* 47(28):7523-7538.

10. Lanz ND & Booker SJ (2012) Identification and function of auxiliary iron-sulfur clusters in radical SAM enzymes. *Biochim Biophys Acta* 1824(11):1196-1212.

11. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, & Miller NE (2001) 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):1097-1106.

12. Hiscox MJ, Driesener RC, & Roach PL (2012) Enzyme catalyzed formation of radicals from S-adenosylmethionine and inhibition of enzyme activity by the cleavage products. *Biochim Biophys Acta* 1824(11):1165-1177.

13. Walsby CJ, Ortillo D, Broderick WE, Broderick JB, & Hoffman BM (2002) 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):11270-11271.

14. Yokoyama K, Ohmori D, Kudo F, & Eguchi T (2008) Mechanistic study on the reaction of a radical SAM dehydrogenase BtrN by electron paramagnetic resonance spectroscopy. *Biochemistry* 47(34):8950-8960.

15. Grove TL, Ahlum JH, Sharma P, Krebs C, & Booker SJ (2010) A consensus mechanism for Radical SAM-dependent dehydrogenation? BtrN contains two [4Fe-4S] clusters. *Biochemistry* 49(18):3783-3785.

16. Benjdia A, et al. (2010) Anaerobic sulfatase-maturating enzyme--a mechanistic link with glycyl radical-activating enzymes? *FEBS J* 277(8):1906-1920.

17. Goldman PJ, *et al.* (2013) X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. *Proc Natl Acad Sci USA (In press)*:10.1073/pnas.1302417110.

18. Haft DH & Basu MK (2011) Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol* 193(11):2745-2755.

19. Fluhe L, *et al.* (2012) The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A. *Nat Chem Biol* 8(4):350-357.

20. Wecksler SR, *et al.* (2009) Pyrroloquinoline quinone biogenesis: demonstration that PqqE from Klebsiella pneumoniae is a radical S-adenosyl-L-methionine enzyme. *Biochemistry* 48(42):10151-10161.

21. Haft DH (2011) Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* 12:21.

22. Mehta AP, *et al.* (2013) Catalysis of a new ribose carbon-insertion reaction by the molybdenum cofactor biosynthetic enzyme MoaA. *Biochemistry* 52(7):1134-1136.

23. Hanzelmann P & Schindelin H (2004) Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc Natl Acad Sci USA* 101(35):12870-12875.

24. Hover BM, Loksztejn A, Ribeiro AA, & Yokoyama K (2013) Identification of a cyclic nucleotide as a cryptic intermediate in molybdenum cofactor biosynthesis. *J Am Chem Soc* 135(18):7019-7032.

25. Hanzelmann P & Schindelin H (2006) 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 USA* 103(18):6829-6834.

26. Lees NS, *et al.* (2009) ENDOR spectroscopy shows that guanine N1 binds to [4Fe-4S] cluster II of the S-adenosylmethionine-dependent enzyme moaA: Mechanistic implications. *J Am Chem Soc* 131(26):9184-9185.

27. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.

28. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.

29. Layer G, Moser J, Heinz DW, Jahn D, & Schubert WD (2003) Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J* 22(23):6214-6224.

30. Nicolet Y, Amara P, Mouesca JM, & Fontecilla-Camps JC (2009) Unexpected electron transfer mechanism upon AdoMet cleavage in radical SAM proteins. *Proc Natl Acad Sci USA* 106(35):14867-14871.

31. Quitterer F, List A, Eisenreich W, Bacher A, & Groll M (2012) Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis. *Angew Chem Int Ed Engl* 51(6):1339-1342.

32. Hayon E & Simic M (1974) Acid-base properties of free-radicals in solution. *Accounts Chem Res* 7(4):114-121.

33. Guillen Schlippe YV & Hedstrom L (2005) A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* 433(1):266-278.

34. Ruszczycky MW, Choi SH, & Liu HW (2013) EPR-kinetic isotope effect study of the mechanism of radical-mediated dehydrogenation of an alcohol by the radical SAM enzyme DesII. *Proc Natl Acad Sci USA* 110(6):2088-2093.

35. Moser CC, Anderson JL, & Dutton PL (2010) Guidelines for tunneling in enzymes. *Biochim Biophys Acta* 1797(9):1573-1586.

36. Ruszczycky MW, Choi SH, & Liu HW (2010) Stoichiometry of the redox neutral deamination and oxidative dehydrogenation reactions catalyzed by the radical SAM enzyme DesII. *J Am Chem Soc* 132(7):2359-2369.

37. Ruszczycky MW, Choi SH, Mansoorabadi SO, & Liu HW (2011) Mechanistic studies of the radical S-adenosyl-L-methionine enzyme DesII: EPR characterization of a radical intermediate generated during its catalyzed dehydrogenation of TDP-D-quinovose. *J Am Chem Soc* 133(19):7292-7295.

38. Szu PH, Ruszczycky MW, Choi SH, Yan F, & Liu HW (2009) Characterization and mechanistic studies of DesII: A radical S-adenosyl-L-methionine enzyme involved in the biosynthesis of TDP-D-desosamine. *J Am Chem Soc* 131(39):14030-14042.

39. Szu PH, He X, Zhao L, & Liu HW (2005) Biosynthesis of TDP-D-desosamine: identification of a strategy for C4 deoxygenation. *Angew Chem Int Ed Engl* 44(41):6742-6746.

40. Ruszczycky MW, Ogasawara Y, & Liu HW (2012) Radical SAM enzymes in the biosynthesis of sugar-containing natural products. *Biochim Biophys Acta* 1824(11):1231-1244.

41. Fluhe L, *et al.* (2013) Two [4Fe-4S] clusters containing radical SAM enzyme SkfB catalyze thioether bond formation during the maturation of the Sporulation Killing Factor. *J Am Chem Soc* 135(3):959-962.

42. Iwig DF & Booker SJ (2004) Insight into the polar reactivity of the onium chalcogen analogues of S-adenosyl-L-methionine. *Biochemistry* 43(42):13496-13509.

43. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.

44. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

45. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213-221.

46. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.

47. Pettersen EF, *et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comp Chem* 25(13):1605-1612.

48. Smoot ME, Ono K, Ruscheinski J, Wang PL, & Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27(3):431-432.

Chapter 5.

Lessons from Recent Structures of AdoMet Radical Enzymes

V.I   SUMMARY

In our investigations of AdoMet radical dehydrogenases, we have come across two very important features of the AdoMet radical fold. One is the SPASM/twitch architecture for binding auxiliary iron sulfur clusters in AdoMet radical enzymes. The prevalence of these enzymes, which are distinct from other auxiliary cluster containing AdoMet radical enzymes, suggests that the feature will be widely represented in the discovery of more and more AdoMet radical enzymes in secondary metabolism. This architecture and its use in predicting AdoMet radical mechanisms are discussed in the first section of this chapter.

Secondly, the enzyme BtrN presents a large deviation in the traditional $(\beta/\alpha)_6$ partial TIM barrel fold thought to be conserved in the AdoMet radical enzymes. Another enzyme recently characterized in the Drennan group, QueE, represents another departure from the core fold. The abridged nature of the AdoMet radical fold in these enzymes helps us understand how AdoMet radical enzymes that seem to be too short to support the $(\beta/\alpha)_6$ fold can support AdoMet radical chemistry. These results and predictions are discussed in the second section of this chapter.

V.II   Auxiliary clusters in AdoMet radical proteins

With the addition of the dehydrogenases, a significant number of auxiliary cluster containing AdoMet radical enzymes have been structurally characterized. These auxiliary clusters carry out a variety of functions (Figure V.1) (1). In a sulfur insertion reaction, BioB (containing an auxiliary [2Fe-2S] cluster) sacrificially donates a sulfur atom from its auxiliary cluster to substrate in the formation of biotin (2-4). Similarly, LipA contains an auxiliary [4Fe-4S] cluster, providing two sulfur atoms to form its vitamin product, lipoic acid, also in a sacrificial manner (5-7). Methylthioltransferases, as their name suggests, transfer a methylthiol group to macromolecular substrates including the S12 ribosomal protein (catalyzed by RimO) and tRNA-A37 (catalyzed by MiaB) (8, 9). Recently it has been shown that these enzymes do not use a sacrificial mechanism, instead they use their auxiliary [4Fe-4S] cluster to directly ligate a yet uncharacterized sulfur containing substrate for transfer to their macromolecular cosubstrate (10). Other auxiliary cluster containing AdoMet radical enzymes include glycyl radical activating enzymes (Hpd-AE, Gdh-AE, and BssD) and hydrogenase cofactor maturating enzymes (HydE and HydG) (1). Our understanding of this latter set of enzymes lacks either the structural or mechanistic information for a detailed analysis of auxiliary cluster function. Another subclass, which we have termed the SPASM/twitch subclass, harbors Aux clusters in

very similar structural folds while catalyzing many different reactions. This subclass is the main focus of Chapters 3 and 4.

Contained in this subclass are the SPASM proteins, encompassing an AdoMet radical subclass that is thought to modify ribosomally translated peptides. Characterization of SPASM containing operons could thus lead to the identification of novel natural products (11). The subclass is named for its biochemically characterized founding members, AlbA (12, 13), PqqE (14), the anaerobic sulfatase maturating enzymes (anSMEs) (15, 16), and MtfC (17), involved in subtilosin A, PQQ cofactor, anaerobic sulfatase, and mycofacticin maturation, respectively. Based on work done by Benjdia et al. (16), the SPASM subfamily was defined by Haft and Basu (11) as 281 AdoMet radical enzymes with a C-terminal $CX_{9-15}GX_4C$—gap—$CX_2CX_5CX_3C$—gap—C motif and assigned the accession TIGR04085. Because of this sequence conservation, the subfamily clusters together in representative diagrams of the AdoMet radical superfamily (Figure V.2). Members of this subfamily are thought to use their 7-cysteine motif to bind two auxiliary (Aux) [4Fe-4S] clusters, one fully protein ligated and one with an open coordination site (11). Based on the redox chemistry performed by the biochemically characterized SPASM members, this open coordination site is hypothesized to ligate substrate to aide in transfer of electrons to or from peptide substrate intermediates.

As discussed in Chapter 3 (18), we have shown that direct ligation is not a function of the Aux clusters in the SPASM protein anSMEcpe. AnSMEs contain a conserved cysteine upstream of the SPASM domains that provide an additional ligand to the active site Aux [4Fe-4S] cluster, resulting in both Aux clusters having full protein ligation (both in the presence and absence of substrate). This result, however, does not preclude other SPASM members from a direct ligation function, especially as AlbA and others do not contain an eighth cysteine in their C-terminal Aux cluster domains. It will be very interesting to discover whether these enzymes i) do indeed have an open Fe coordination site for substrate ligation, ii) if they use another amino acid for ligation, as in BioB, which has an arginine ligand (19), or iii) if they use a cysteine that is distal in primary sequence from its SPASM motif to provide the last Aux cluster ligand.

In our structural characterization of the 7-cystiene motif in anSMEcpe, a few surprising results emerged. First, the active site Aux cluster (Aux I) is ligated by three cysteines surrounding a beta hairpin motif. This beta hairpin extends the N-terminal end of the AdoMet radical partial TIM barrel (see Chapter 1). Follow these ligation sites and a helical region that seals Aux I from solvent, the $CX_2CX_5CX_3C$ provides three cysteine ligands to the second Aux cluster (Aux II). Interestingly, the fourth cysteine of the $CX_2CX_5CX_3C$ motif crosses back over the domain to provide the final Aux I ligand. After this site, the final cysteine of the SPASM domain

ligates Aux II. Because of the sequence homology of the SPASM domain members (see Chapter 3), we believe that this intricate cluster binding architecture will be a conserved feature of SPASM members. Also, the first iron ligating cysteine in anSMEs, the 'non-SPASM' cysteine, is a ligand to Aux I, the active site Aux cluster. If future members of the SPASM subclass do indeed use an Aux cluster for substrate ligation, the anSMEcpe architecture is consistent with this ligation site being proximal to the active site.

We have also characterized the structure of BtrN (Chapter 4). While this enzyme catalyzes the same chemistry as anSMEcpe (Figure V.1) (20, 21), it harbors only one Aux cluster and so, is not a member of the SPASM subclass (22). Sequence clustering reflects this difference and places BtrN on the outskirts of a large cluster comprised of BtrN, MoaA (see below), and the SPASM proteins. This sequence cluster is separate from the other Aux cluster containing AdoMet radical enzymes outlined above. We find that structural elements surrounding active site Aux clusters in the anSMEcpe and BtrN structures overlay very well (RMSD 1.1 Å). These structurally similar elements include the beta-hairpin that sits between cysteine ligand sites and the alpha helical region, which precedes the $CX_2CX_5CX_3C$ motif in anSMEcpe. Generally, sequence conservation is limited in this region, except for the beta hairpin motif itself, where a conserved glycine sits in the i+3 position of the turn. This Gly residue is related to [4Fe-4S] ligating cysteines by a $CX_{9-15}GX_4C$ motif that is strictly conserved in the SPASM subfamily (see Figure III.2).

MoaA, another structurally characterized AdoMet radical enzyme, also contains this beta turn around an Aux cluster. This protein has some sequence homology to anSMEcpe and BtrN and like BtrN, MoaA only contains one Aux cluster. It houses this cluster using the same beta-hairpin and alpha helical region as in BtrN and anSMEcpe, but does not have any iron ligating cysteines after the domain (23). As MoaA and BtrN share the half-SPASM motifs out lined above, we have classified their Aux cluster binding domains as 'twitch' domains. MoaA is also part of the sequence cluster containing BtrN and anSMEcpe, so we have designated this large sequence cluster, the SPASM/twitch cluster (Figure V.2). Interestingly, MoaA has only three protein ligating cysteines, using the unique iron position on its Aux cluster to ligate substrate (24), giving precedence to the SPASM/twitch enzymes using Aux clusters for direct substrate ligation. MoaA also catalyzes a redox reaction (25, 26).

145

We now have a set of three AdoMet radical enzyme structures with Aux clusters bound in SPASM/twitch domains. We also have a tool in the Clustal mapping of the AdoMet radical superfamily, which correlates well with our structural data. While the sequence mapping contains a broad set of enzymes, we have structural data from three edges of the SPASM/twitch cluster (Figure V.2). Because of this sequence disparity, we believe our set of SPASM/twitch enzymes will be among the most diverse in the cluster. Despite this, however, the Aux cluster domains of the three enzymes show a high degree of structural conservation (Chapter 4), leading us to believe the Aux cluster folds of BtrN, anSMEcpe, and MoaA will be a feature of a majority of SPASM/twitch proteins. Of the ~ 3,500 AdoMet radical superfamily nodes, which contain unique sequences with ≥ 40% sequence identity, the SPASM/twitch cluster accounts for 1,063. Hence, we believe that the SPASM/twitch enzymes will comprise a large segment of the AdoMet radical diversity. Consistent with this hypothesis, all characterized SPASM/twitch enzymes are involved in secondary metabolism. We have only begun scratching the surface of this interesting subclass.

One area of recent focus in this family is the thioether bond formation family. AlbA (12) and SkfB (27), members of this family, catalyze identical linkages of a cysteine to an up or downstream Cα position in the maturation of the subtilosin A and sporulation killing factor, two natural products generated from ribosomally translated peptide scaffolds (Figure V.1). Following hydrogen abstraction from the Cα position, formation of a covalent bond with this cysteine necessitates a one electron oxidation and deprotonation of the sulfur position. The parallels in this system to the dehydrogenases (deprotonation and one electron oxidation) again made direct ligation an attractive hypothesis (12, 27). AlbA contains seven cysteines in a SPASM motif (and is a founding member of the SPASM subfamily). SkfB, however, only contains five cysteines in its C-terminal domain. An alignment with the structurally characterized members of the SPASM/twitch subclass suggests that SkfB will contain a twitch domain like that of MoaA and BtrN (Figure IV.10). These analyses suggest that, like the AdoMet radical dehydrogenases, the thioether bond forming enzymes will use both the SPASM and twitch architectures to catalyze identical reactions.

Using the structures of BtrN, anSMEcpe, and MoaA, we are in a position to comment on this hypothesis from a structural perspective. Because of the conserved nature of AdoMet cluster – Aux cluster distances, we can expect the distance between the two clusters to be ~ 16 Å in both AlbA and SkfB. Using these three examples of SPASM/twitch domains we can further estimate the distance from hydrogen abstraction site to Aux cluster to be ~ 9.4 Å (an average of

9.6, 8.6, and 10.1 Å in BtrN, anSMEcpe, and MoaA, respectively) (Figure V.3). In AlbA and SkfB, after subtracting for a typical S-Fe bond distance, we are left with ~ 7 Å for the distance from the Cα hydrogen abstraction site to the directly ligated cysteinyl sulfur position. In MoaA, where a molecule of GTP is both the hydrogen abstraction site (2' hydroxyl) (25) and the Aux cluster ligation site (N1 of guanosine base) (28), this separation is both intramolecular and distal (8.1 Å, Figure V.3) (24). Manipulation of this arrangement to bring the hydrogen abstraction site directly between the 5' position of AdoMet and the Aux cluster, only decreases this distance by ~ 1 Å (Figure V.3). Without an unprecedented conformational change accompanying AdoMet radical activity, it would seem that a cysteine ligand directly bound to an Aux cluster would be too far from a Cα hydrogen abstraction site to participate in chemistry. If further studies find that AlbA, SkfB, or any other SPASM/twitch family protein do indeed use a direct ligation model, we hypothesize that this function will only be an anchoring one; any directly bound ligand seems to be structurally prohibited from participating in chemistry initiated at the AdoMet cluster.

We can contrast these predictions with auxiliary cluster containing AdoMet radical enzymes in different families where auxiliary clusters are proven to be involved in chemistry. For example, in the sulfur insertion enzyme BioB, a dethiobiotin radical intermediate acquires a sulfur atom from the bridging position of its auxiliary [2Fe-2S] cluster (29), which is ligated by three cysteines residing in the middle of β2, β3, and β5 strands and an arginine that sits outside of the AdoMet radical fold (19). The distance between this hydrogen abstraction site and the Aux cluster in this system is 4.6 Å, ~ 5 Å closer than in the SPASM/twitch architecture (Figure V.3). The structure of RimO, a methylthioltransferase (MTT), was also recently solved. Along with presenting the structure of RimO, the authors determined that the sulfur source in MTTs is not the Aux cluster. The sulfur donor remains undetermined, but is suggested to bind at an available coordination site on the 3-cysteine ligated Aux cluster, as hypothesized for the thioether bond forming enzymes (10, 12, 27). In RimO, however, the FeS clusters are 8.4 Å apart, which is even closer than in BioB (12.0 Å) and nearly twice as close as in the SPASM/twitch architecture (the RimO structure does not contain AdoMet or cosubstrate) (Figure V.3). The structures of BioB and RimO display AdoMet and Aux clusters positions that agree with studies suggesting that these Aux clusters play a direct role in chemistry. The distances between the two clusters are much smaller than distances observed in the SPASM/twitch fold, consistent with the idea that while some AdoMet radical enzymes employ auxiliary clusters in catalysis, cluster separation in the SPASM/twitch architecture appears too great for this function.

V.III Abridged AdoMet folds

With few exceptions, the core partial TIM barrel architecture of AdoMet radical fold is consistent among the structurally characterized AdoMet radical enzymes (see Chapter 1) (30, 31). Structural motifs for AdoMet cluster and AdoMet binding are consistent throughout the family and allow us to expect certain features of the AdoMet radical when solving new structures. Interestingly, these motifs have very little sequence similarity between family members, as they rely primarily on peptide backbone amides. The exception to this backbone based interaction pattern has traditionally been in the CX_3CX_2C motif itself, a commonality among AdoMet radical enzymes that led to the initial identification of the superfamily (32). What we thought we knew about AdoMet radical folds has been put to the test recently, especially in the characterization of BtrN (Chapter 4) and 7-carboxy-7-deazaguanine synthase, QueE, from *Burkholderia multivorands* (Daniel Dowling, unpublished). These two enzymes have provided two distinct exceptions to the AdoMet radical fold, both of them involving abridged AdoMet radical domains that are able to initiate AdoMet radical chemistry in the same manner as their more traditional AdoMet radical cousins. Additionally, QueE has expanded our understanding of the very nature of the AdoMet radical CX_3CX_2C sequence motif itself, presenting the possibility that the AdoMet radical superfamily is much larger than previously thought.

The $(\beta/\alpha)_6$ partial TIM barrel fold thought to be conserved in AdoMet radical proteins, uses 215 ± 12 residues to complete the fold (from the first residue of $\beta1$ to the last residue of $\alpha6$) (Figure V.4). Some AdoMet radical proteins contain this partial barrel within the context of a full $(\beta/\alpha)_8$ TIM barrel, while the majority use additional structural elements or substrate binding partners to seal the interior of the partial barrel from solvent. Prior to this work, however, all characterized AdoMet radical protein structures had at least six β strands and six α helices. Now we see that BtrN, as discussed in Chapter 4, lacks 25% of this partial TIM barrel architecture by completing its AdoMet radical domain in a β_5/α_4 partial TIM barrel, omitting the last two helices and a strand from a traditional AdoMet radical fold (Figure V.5.). QueE uses a different tactic. In this enzyme, all six strands of the barrel are conserved, however, instead of using antiparallel helices to connect each strand, helices 3-5 are substituted by loops, requiring far fewer residues to continue the parallel beta sheet core (Figure V.5.). The results of these alterations are that BtrN and QueE complete their AdoMet radical folds in 152 and 162 residues, respectively, compared to the typical 215 ± 12 (Figure V.4).

These abridged AdoMet radical folds preserve a very similar primary coordination sphere around AdoMet by conserving the four AdoMet binding motifs described previously, including the 'GGE' motif, the ribose motif, the 'GXIXGXXE' motif, and the $\beta6$ motif (see Chapters 3 and 4 for detailed descriptions of these motifs). Because none of these motifs are found in helical regions of the partial barrel, the fact that QueE lacks 3 helices of the AdoMet radical fold does not prevent the enzyme from using all four motifs. In BtrN, a small turn after $\beta5$ is able to provide very similar 'GXIXGXXE' and $\beta6$ motif contacts using many fewer residues than required by the traditional AdoMet radical fold (see Chapter 3).

The lack of sequence similarity of the structurally characterized AdoMet radical enzymes makes it difficult to predict other enzymes that will have an abridged AdoMet radical domain. DesII could have an abridged AdoMet radical domain because of its sequence similarity to BtrN (see Chapter 4). In addition, the class III ribonucleotide reductases (RNRs) are also expected to have an abridged fold due the drastically smaller sizes of these enzymes compared to other AdoMet radical enzymes. *E. coli* class III RNR, NrdG, is 156 residues (compared to 250 and 210 for BtrN and QueE, respectively) and was hypothesized to have an $(\beta/\alpha)_4$ half-TIM barrel due it this drastically abridged nature (33). With the structures of BtrN and QueE, our knowledge of the AdoMet radical family has since expanded. With these two structures as reference models in the I-TASSER server (34), we have generated two NrdG models (Figure V.5). Both models contain a $\beta1'$ strand N-terminal to the AdoMet radical fold, due to the similarity of NrdG to QueE and another AdoMet radical enzyme with a similar extension, pyruvate formate lyase activating enzyme (PFL-AE) (35). As the CX_3CX_2C motif starts with C26 in NrdG, there is sequence space for this N-terminal extension. Both NrdG models contain cysteines in a loop following $\beta1$, in a conformation that supports binding a [4Fe-4S] cluster. In addition, both NrdG models adopt traits from BtrN and QueE. First, like BtrN, the AdoMet radical fold of NrdG terminates after the fifth AdoMet radical beta strand (not including the $\beta1'$ strand, colored in grey in Figure V.5).

Additionally, like QueE, the NrdG models are missing helical regions. Instead of a BtrN-like β_5/α_4 fold, the NrdG model has a β_5/α_3 structure with the $\alpha4$ helix substituted by a loop. The AdoMet cluster and AdoMet cofactors fit very well into the NrdG model. Further, the 'GGE' motif (GGD in NrdG, Figure V.5.D, green spheres) and the ribose motif (T105, Figure V.5.D, yellow spheres) can be assigned based on their sequence and positioning in the model. The $\beta5$ extension, which we can predict will contain both the 'GXIXGXXE' and $\beta6$ motif AdoMet interactions (as in BtrN), will require adopting a different fold in a true NrdG structure as their predicted positions (in both BtrN and QueE derived structures) don't to complete the AdoMet

149

binding pocket. It will be interesting to see if the architecture of the loop following β5 in BtrN is conserved in other abridged AdoMet radical proteins. Even with a restructuring of β5, however, the active site of the NrdG model is highly solvent exposed. This result is expected, as the function of the enzyme is the activation of the catalytic subunit of RNR. Binding to this partner protein is thought to seal the active site from solvent, as in the case of PFL-AE binding to its partner protein, pyruvate formate lyase (35).

The structures of QueE and BtrN have expanded the possibilities of AdoMet radical core folds. By being able to complete their AdoMet radical folds in about 60 residues less than other members, BtrN and QueE have provided examples of how enzymes that are smaller than the traditional 215 residue minimum can support AdoMet binding. QueE from *B. multivorands* is also unique in its AdoMet cluster binding motif. Instead of binding this cluster in the traditional CX_3CX_2C motif, one of the few unifying sequence motifs in the family, QueE binds its AdoMet cluster in a $CX_{14}CX_2C$ (Figure V.4). This insertion between the first two iron ligating cysteines forms a small helix above the AdoMet binding loop, which leaves the core of the AdoMet fold intact (Figure V.5). QueE from *Bascillus subtilus* (36) does not contain this insertion (Figure V.4), suggesting that the helix is not involved in the QueE mechanism. While the function of the helix is still undetermined, its existence raises the possibility of large numbers of AdoMet radical enzymes that have not yet been identified because their AdoMet cluster binding sequences deviate from the expected CX_3CX_2C motif.

V.IV   CONCLUSION

The enzymes anSMEcpe and BtrN use their [4Fe-4S] cofactors both to bind a substrate (AdoMet) and to facilitate electron transfer. While we have disproved the hypothesis of direct substrate ligation to an Aux cluster in the dehydrogenases, we already have an example in MoaA of a SPASM/twitch enzyme that has this feature. Because of the diversity in the SPASM/twitch cluster, we believe there is a high likelihood that another protein in this sequence cluster uses an available coordination site on an Aux cluster for direct ligation to substrate. However, as in MoaA, we do not believe this feature will participate in chemistry, as was hypothesized for the dehydrogenases and remains hypothesized for the thioether bond forming enzymes. This conclusion is mainly due to the large difference in AdoMet – Aux cluster distances in the SPASM/twitch architecture compared to systems proven to use their Aux clusters in catalysis.

Recently, we have also seen exceptions to the AdoMet radical partial TIM barrel core fold. It seems that these exceptions will be present not only in AdoMet radical enzymes whose sequences are too short to include all facets of the AdoMet radical fold, but larger enzymes, such as BtrN and QueE, in which predicting these abridged features becomes much more difficult. As more structures of AdoMet radical enzymes become available, the SPASM/twitch Aux cluster architecture and modifications on the core TIM barrel fold hypotheses will be put to the test. Either way, the rate at which these structures are becoming available will make the upcoming years a fascinating time in the AdoMet radical structure field.

Figure V.1. AdoMet radical enzyme reactions requiring auxiliary cluster(s).

subtilosin A / sporulation killing factor
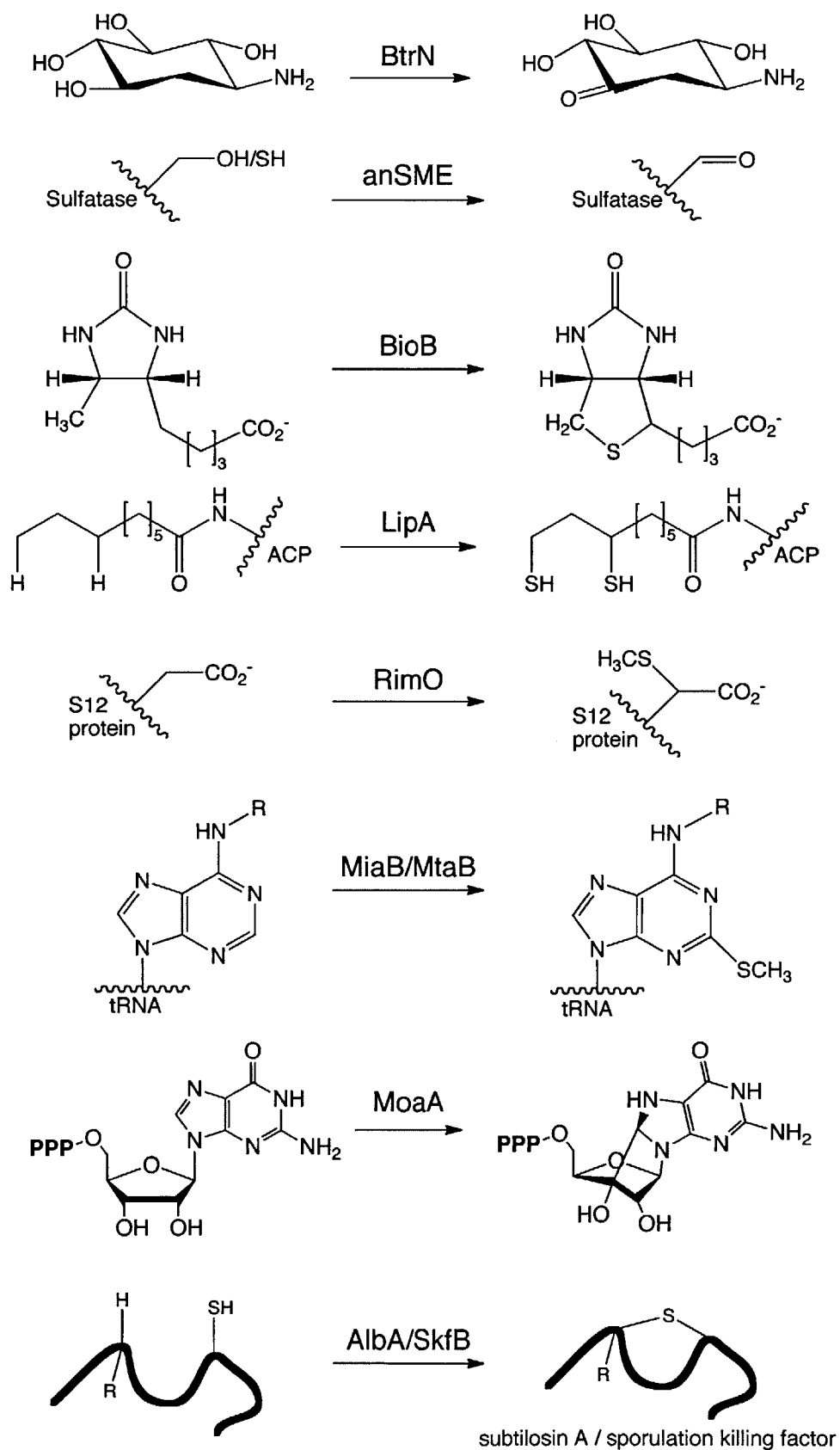
Figure V.2. Sequence space diagram showing the relationship of Aux cluster containing AdoMet radical enzymes in the superfamily. Sequence mapping of the auxiliary cluster contain enzymes discussed in this chapter and Figure 1, generated using an $E = 10^{-22}$ E-value cutoff in Clustal (37). SPASM enzymes that are not biochemically characterized are in green. For more complete representations of the superfamily, see Figures III.14 and IV.12.
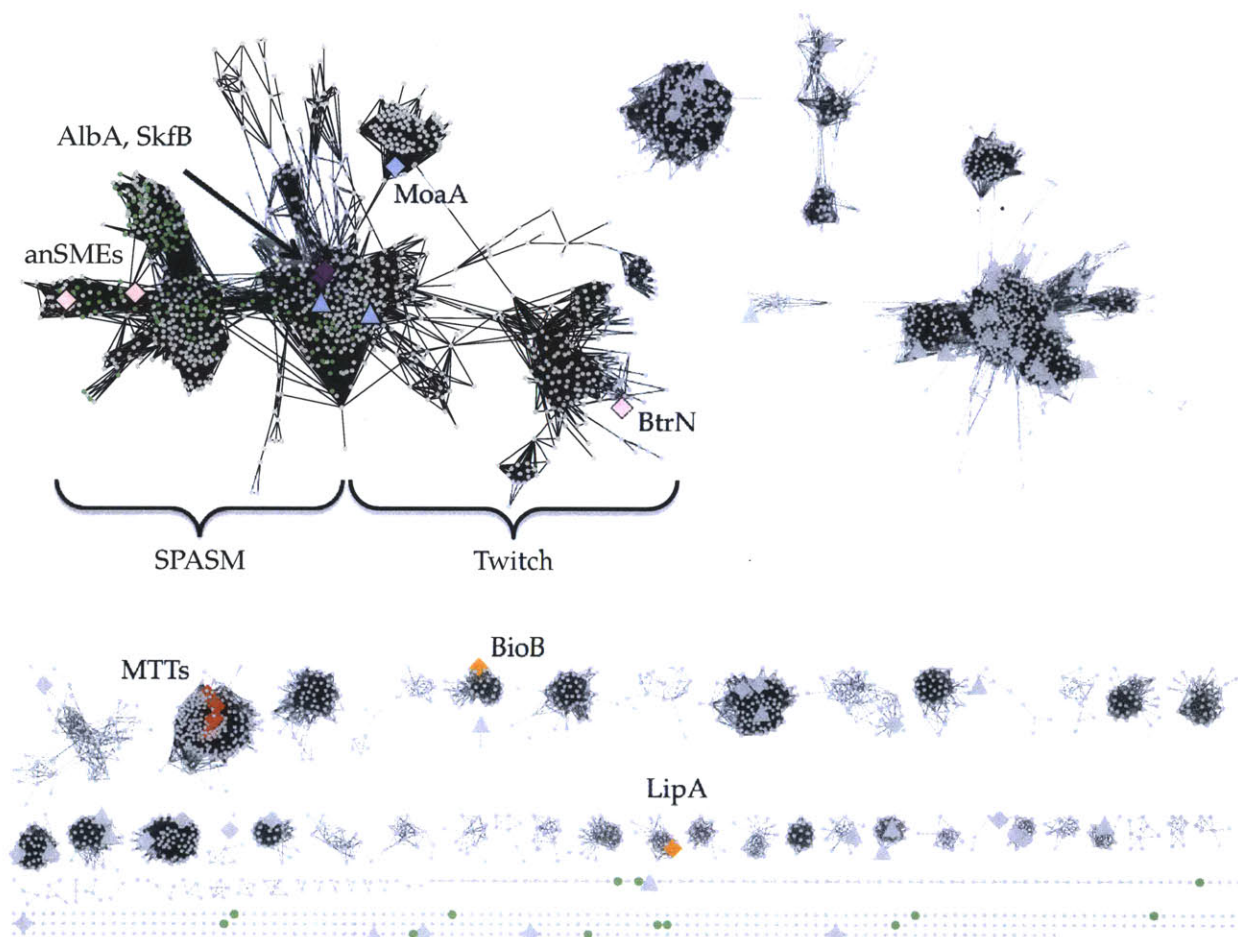
Figure V.3. Auxiliary cluster positions in SPASM/twitch enzymes. Hydrogen abstraction position to Aux cluster distances in (*A*) BtrN (9.6 Å), (*B*) anSMEcpe (8.6 Å), and (*C*) MoaA (10.1 Å) are indicated by black dashes. (*D*) Model of AlbA active site where the distance from hydrogen abstraction site (Cα carbon of a glycine residue) to an Aux cluster bound cysteine residue is indicated by red dashes (6 Å in the model). In the non SPASM/twitch enzymes (*E*) BioB and (*F*) RimO the two FeS clusters are much closer (AdoMet is modeled into the RimO structure, which lacks both AdoMet and substrate). AdoMet to Aux clusters distances in BtrN, anSMEcpe, MoaA, BioB, and RimO are 15.8, 16.8, 16.3, 12.0, and 8.4 Å, respectively.

Figure V.4. Structure based sequence alignment of AdoMet radical folds. Strands in yellow, helices in cyan, AdoMet binding cluster cysteines in red, residues interacting with AdoMet in green, residues interacting with substrate in blue, residues interacting with both AdoMet and substrate in brown, residues interacting with an additional cofactor (i.e. an FeS cluster) in pink.

```
                    --β1--
PFLA_EC   -SVIGRIHSFESCGTVDGPGIRFITFF-----QGGLMR----------LMHNRDTWDTHGGKEVTV-------------------- 72
HEMN_EC   SEDFGEQAFLQAVARYPERPLSLYVHI-----PFTHKL----------VEGCNKIVTRQQ-----------------------HK 83
BIOB_EC   LLFEAQQVHRQHFDPRQVQVSTLLSIK----TGAPED----------KYPQSSRYKTGLEAERLM-------------------- 68
MOAA_SA   --------MVEQIKDKLGRPIRDLRLS--VTDRNFR----------DYMPKEVFGDDFVFLPKNELLT----------------- 51
KAMA_CS   DPLHEDTDSPVPGLTHRYPDRVLLLIT-----DMSMY----------RHLTRRFAGQSDDSMP--------------------- 146
TYW1_PH   KKSLTEGRFCYKQKFYGIHSHRCLQMTPVL--AWTHN----------IEWRPMENFLGTELPQP----WDDPAFIVEESIKAQRKLL 122
TYW1_MJ   RKKMLEDKNCYKSKFYGIETHRCIQCTPSV--IWQQN----------IEWRVLPRDIGIDISQIKEPKWEEPEVVVYEKILAMHKRII 107
HYDE_TM   LFKLADEIRRKYVGDE-VHIRAIIEFS----NVRKN----------LYIGLRRDNKN--LKRYRM--------------------- 84
RLMN_EC   ------VETVYIPEDD-----RATLCV--SSQVGALE----------KFFSTAQQGFNRN--LR------------VSEIIGQ 103
RIMO_TM   --------------EERPYAYVK---ISDGDRG----------TESSIPSFKGSLRSRS----------------------- 168
SPL__GT   ---------LKFDSSKPSAEYAIPLAT------GMGH----------VYLQTTLGSKPYIRVYVN-------------------- 114
ANSMECP   ----------------MPPLSLLIKPASS-GNLK----------TYHSLSDNRNVKSYG-------------------- 37
BTRN_BC   ---------------MDKLFSMILVEVNS--GNRT----------WYIPNSVSKRKETGEMD------------------ 37
PYLB_MB   ---------FGNRVFLNCFIYFST----YSKNQ----------SFYYNCRNEINRYLT-------------------- 90
QUEE_BS   MAKGIPVLEIGPSLCGMGVIGQKTMFVTAGDYS----------SWFLAFTWDGSAKKDIRWM------------------- 58
QUEE_BM   --MTYAVKEIPLGANAGRPAVLCLFAGNLWSGREEDRAQAVIGIDFVGTDGENGG-KFKD------------------- 66

              ---α1---              --β2--                    ---α2---               --β3--
PFLA_EC   --EDLMKEVVTYRHFMNAS-G----GGVTAS----G-----GE-----------AILQAEFVRDWFRACK---KE-GI---HTCLVTN------GF 128
HEMN_EC   ADQYLDALEQEIVHRAPLFAGRHVSQLHWG----A------GTPT----------YLNKAQISRLMKLLRENFQFN-AD--AEISIBVD------PR 151
BIOB_EC   -----EVEQVLESARKAKAAG--STRFSMG----A-----AWKN----------PHERDMPVLEQMVQGVKA----MG---LEAGMTL------GT 125
MOAA_SA   -----FDEMARIAKVYAELG--VFKILIT----G-----GE-----------PLMRRDLDVLIAKLNQID-GI--EDIGLGTN------GL 106
KAMA_CS   -----MERIDKAIDYIRNTPQ-VRDVLLIG----A------GTPT----------ALLVSDETLEYIIAKLR---EIPH--VEIVKIGSRTPVVLPQ 209
TYW1_PH   IGYKGNP-KVDKKKFEEAWNPT-----HAA-----ISLSGE----------PMLYPYMGDLVEEFHKR-----G--FTTFIVTN------GT 180
TYW1_MJ   MGYAGVLDRVGEKKFKEALEPK-----HVA-----ISLSGE----------PTLYPYLDELIKIFHKN-----G--FTTFVVSN------GI 166
HYDE_TM   -----TPEEIVERARLAVQFG--AKTIVLG----T-----GE-----------DPYYMPDVISDIVKEIKK---MG----VAVTLSL------GV 139
RLMN_EC   VWRAAKIVGAAKVTGQRPI-T----NVVMM----GM-----GE-----------PLLNLNNVVPAMEIMLDDFGFGLSKRRVTLSTS------GV 167
RIMO_TM   -IEDITREVEDLLKE-----GK-KEIILV----A-----QD--TTSYGIDLYRKQALPDLLRRLNSL----NGEF----WIRVMYLH------P- 230
SPL__GT   ----LDDIFAQAQKYINER-A--PEITRFEAAT----S-D-----------IVGIDHLTHSLKKAIEFIGA--TDY--GRLRFVTTY------- 175
ANSMECP   -IMRDEVLESMVKRVLNEANG-----HCSFAF---G-----GE-----------PTLAGLEFFEKLMELQRKHN--YKNLKIYNSLQTN------GT 102
BTRN      ------PALYKTLMEQLSSLDFA--GRISFT-FY----GE-----------PLLCKNLDLVFGMTTEYI--PRA-----RPIIYTN------GD 94
PYLB_MB   --------MEEIKETCKTLKGAGFHMVLLT----M--GE-----------DPYYYEDPNRFVELVQIVKEELG----LPIMILP-------G 148
QUEE_BS   ------TAEEIFAELKDIGGD---AFSHVTI---L-----GN-----------PALLKQLDAFIELLKENN-----IRAALETQG------ 111
QUEE_BM   ------ADALVATIAGLWPAGEAHRFVVCL---G-----GE-----------PMLQLD---QPLVDALHAAG-------FGIAIETNGSLP---- 122

              --α3--        --β4--      -α4a--              -α4-          --β5--
PFLA_EC   VRRYDPVIDELLE--VTD-LVMLDLR--QMNDEIHQNLV-GVSN----HRTLEFAKYLANKNVK---VWIRYTV-VPGWSDDDD-----SA 200
HEMN_EC   EIELDVLDHLRAE--GFN-RLSMGVQDFNKEVQRLVNGEQ----DEEFIFALLNHAREI--GFTST---NIDLYGLPKQTPE-----SFAF 226
BIOB_EC   -LSESQAQRLANA--GLD-YYHNLGTS--PEFYGNIIT--TRT--YQERLDTLEKVRDA-GI---KVGSGGLETVK------DRAG 196
MOAA_SA   LLKKHGQKLYDA---GL-RRIIVSLGAIDDTLFQSINNR-N-I--KATTILEQIDYATSI-GL---NVKVTVVIQKGINDD-----QIIPM 180
KAMA_CS   RITPELVNMLKKYH-P--VWLNTHFNH-----------PNEIT---EESTRACQLLADA-GVP---LGNQSVLLRGVNDCVH-----VMK 273
TYW1_PH   -IPERLEEMIKEDKLPT--QLYVSIT-APDIETYNSVMIPMIP----SFENRLRFLELMRDL-PTRTVVRLBL-VKGENMHSP------EK 256
TYW1_MJ   -LTDVIEKIE-----PT--QLYISLD-AYDLDSYRRICGG----KK-EYWESILNTLDILKEK--KRTCIPTTL-IRGYNDDI-----LKF 235
HYDE_TM   -WPREYYEKWEA---GAD-RYILRHE-TANPVLHRKLRPDT---SFENRLNXLLTLKEI-GY---ETGAGSMV-GLPGQ-----TIDDLV 211
RLMN_EC   VPA-LDKLGD-MI--DVA--LAISLH--APNDEIRDEIVPINKKYNIETFLAVRNVSNANQGRVTIEYVM-LDHVNDG-----TEHA 244
RIMO_TM   DHLTEEIISAMLELDKVVKYFTVPVQ--HGSDKILKLMG-RTKS---SEELKKMLSSIRERFPDAV----LRTSI-IVGFPGETEEDFEELK 311
SPL__GT   --EHVDHLLDARHN-GK-TRFTFSIN----SRYVINHFEPGT---SSFDGRLAAARKVAGAGY---KLGFYVAP-IYRHEG----WERGYFE 248
ANSMECP   LIDESWAKFLSE---NK-FLVGLSMG--GPKEIHNLNRKDCCGLDTFSKVERAAELFKKY-KV----FNILCVV-TSNTARH---VNKVYK 179
BTRN      FLTEKRLQTLTEL--GI-QKFIVTQRAGAK-------KFRGVYDQLAGADKE----KVVYLDHSD-------------- 147
PYLB_MB   LMDNATLLKAREK--GA-NFLALYQTYD-TELYRLRV-GQS--FDGRVNARRFAKQQ-GY--CVEDGILT-GVGND-------IESTI 220
QUEE_BS   ---TVYQDWFTLID------DLTIGPKPPSSKMVT------NFQKLDHILTSLQENDRQHA--VSLRVVI---------FND--EDL 170
QUEE_BM   ------------VLESIDWICVSPRADAPLVVT------------------------KGNELKVVI---------PQDNQRL 159

              --α5--       --β6--           ------α6------
PFLA_EC   HRLGEFTRDMG----NVEKIELLPYHELGKHZWVAMGEEYKLDGVKPPKKETMERVKGILEQYGHKVMF-------------------- 265
HEMN_EC   TLKRVAELNFD-------RLSVENYAHLPTIFAAQRKIKDADLPSPQQKLDILQETIAFLTQSGYQFIGMDHFARPDDELAVAQREGVLHRN 311
BIOB_EC   LLLQLANLPTP--PE--SVPIKMLKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVELSAGREQMNEQTQAMCFMAGANSIFYGCKHI 283
MOAA_SA   LEYFKDKHIE-------IFIEFDVGNDNGWDFSKVVTKDEMLTMIEQHFEIDPVEPYFGEVAKYYRHKDNGVQFGLTSVSQSFGST 263
KAMA_CS   ELVNKLVKIRVRP-----PIKQCLSLGLEHFRTPVSKGIEIIEGLRGHTSYCVPTFVVLAPGGGGTPVMPNYVISQSHDKVILRNFE 359
TYW1_PH   YAKLILKARPM------FVEARAYMFVGYSRNRLTINNMPSHQDIREFAEALVKHLPGYHIDEDEYEPSRVVLIMRDDVDPQGTGVEGRFI 340
TYW1_MJ   VELYERADVH-------FIELKSYMHVGYSQKRLKKEDMLQHDEILKLAKMLDENSSYKLIDDSEDSRVALLQNENRKINPKL------- 311
HYDE_TM   DDLLFLKEHD----FD--MVGIGPFIPHPDTPLANEKKGDFTLTLKMVALTRILLPDSNIPAHFAMGTIVPGGREITLRCGANVIMPNWTP 296
RLMN_EC   HQLAELLKDTP-----CKINLPWNPFPGAPYGRSSNSRIDRFSKVLMSYGFTTIVRKTRGDDIDAACGQLAGDVIDRTKRTLRKRMQGE 329
RIMO_TM   QFV-EEIQ-------FD--KLGAFVYSDEEGTVAFNLKEKVDPEMAKRRQEELLLQAEISNSRLDRFVGKKLKFLVEGKEGKFLVGRTWTE 393
SPL__GT   LFQELARQLEGMDLSDLTFELFQHFFTK-------------PAKRVIEQRYPKTRLDLDETKRKYTWGRYGIGFYVVRDEEAKELEDTMR 325
ANSMECP   YFKEK----------DFKBLRIFNCIDPLYEEKGKYNYSLKPKDYTKFLKNLFDFWYEDFLNGNRVSIKYFDGLLGTILLGKSISGMNGT 260
BTRN      ----------------VLSNGGILDNIPQASKANMSMVPSNLAVVTVLGGNIGEQHISDIWHNDKFTSFR 218
PYLB_MB   LSLRGMSTND----PD-MVGVMTFPQEGTPLEGFRDKSNLSELKIISVLRLMFPKRLIPAILLEGIDGMVLRLNAGANIVFILPPDSQ 306
QUEE_BS   EFAKTVHKRYPGIP-----FYLQVGNDVHTTD------DQSLIAHLLGKYEALVDKVAVDAELNLVRVLPQLLTLLWGNKRGL------- 243
QUEE_BM   ADYAKLDFE--------YFLVQPMDGPS------------------DLNTKLAIDWCKRHPQWRL--SMGTEK--YLN--I-------- 210

PFLA_EC   --------------------------------------------------------------------------------------
HEMN_EC   IQGYTTQGDTDLLGMGVSALSMIGDCYAQNQKELKQYYQQVDEQGNALWRGIALTRDDCIRRDVIKSLICNFRLDYAPIEKQWDLHFADYF 402
BIOB_EC   IPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMTPDTDEYYNAAAL-------------------------------- 338
MOAA_SA   TTIALSSDGKFYGGLFATVDGFNVKAFIRSGVTDEELKEQFKALWQIRDDRYSDERTAQTVANRQRKKINMNYIGG---------- 340
KAMA_CS   GVITTYSEPINYTPGCNCDVCTGKKKVHKVGVAGLLNGEGMALEPVGLERNKRHVQE------------------------------ 416
TYW1_PH   KH------------------------------------------------------------------------------------ 342
TYW1_MJ   --------------------------------------------------------------------------------------
HYDE_TM   SPYRQLYQLKPGKIGVFEKDTAGIPEVMKMIELLGRKPGRDWGGRKRVFETV--------------------------------- 348
RLMN_EC   AIDIKAV------------------------------------------------------------------------------ 336
RIMO_TM   APEVDGVVFVRGKGKIGDFLEVVIKEHDEYDMWGSVI------------------------------------------------- 430
SPL__GT   RYIEQFFPGAYVQFTIPASLDLEGIDG---------------------------------------------------------- 341
ANSMECP   FTCQFVVESDGSVYPGLFYVLDKWRLGNIQDMTMKELFETNKNHEFIKLSFKVHEEGKKGKWFRLGKGGGRRCRDSKEDSALELNYYIQSY 351
BTRN      KMLKEGHRGKSDLGKNNNVSVQTEEQGDYIL------------------------------------------------------- 250
PYLB_MB   LEGVANYDRDLEERDRDIKSVVRRLEIMGMKPARQADFEAVLGC----------------------------------------- 350
QUEE_BS   --------------------------------------------------------------------------------------
QUEE_BM   --------------------------------------------------------------------------------------
```
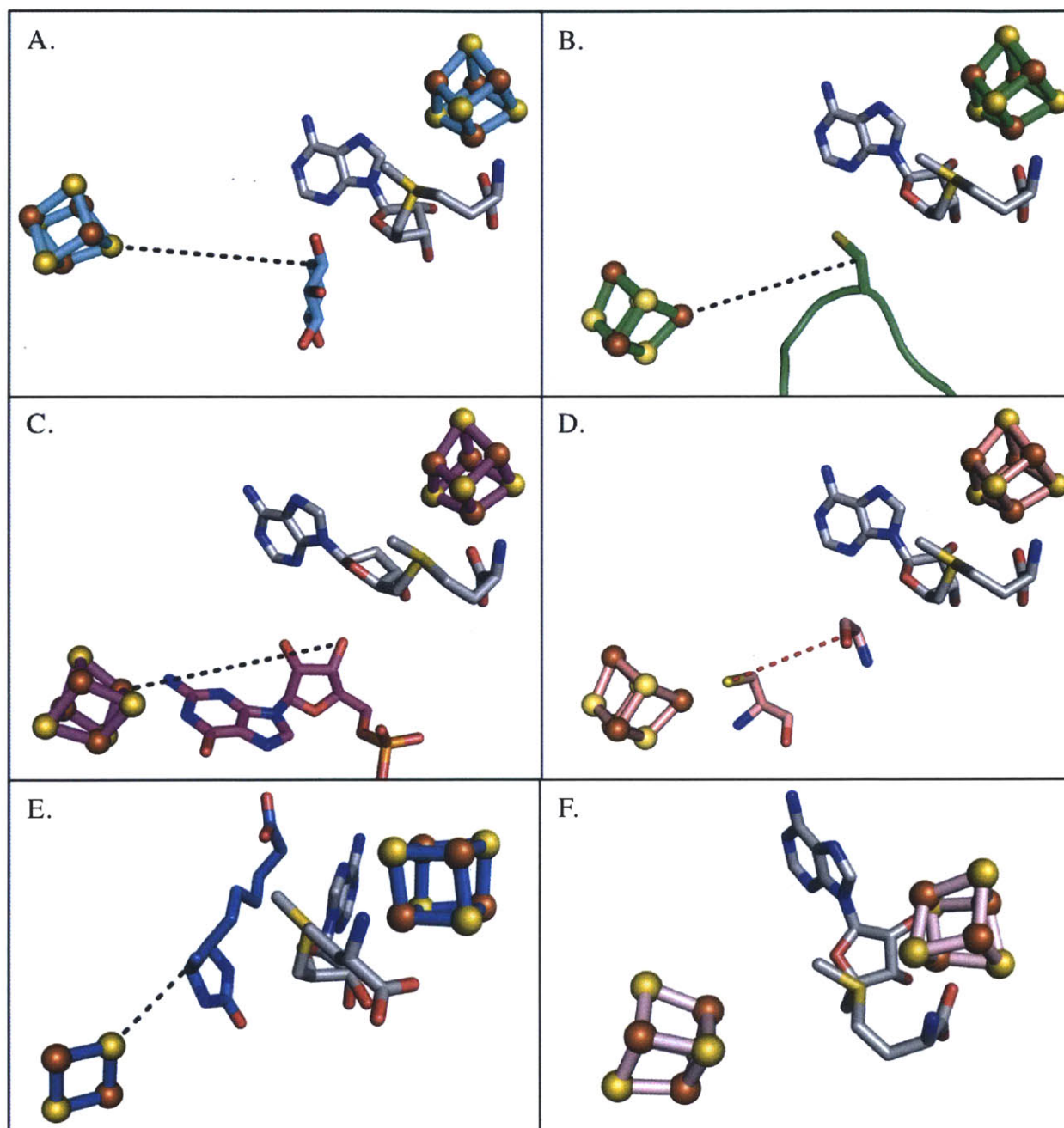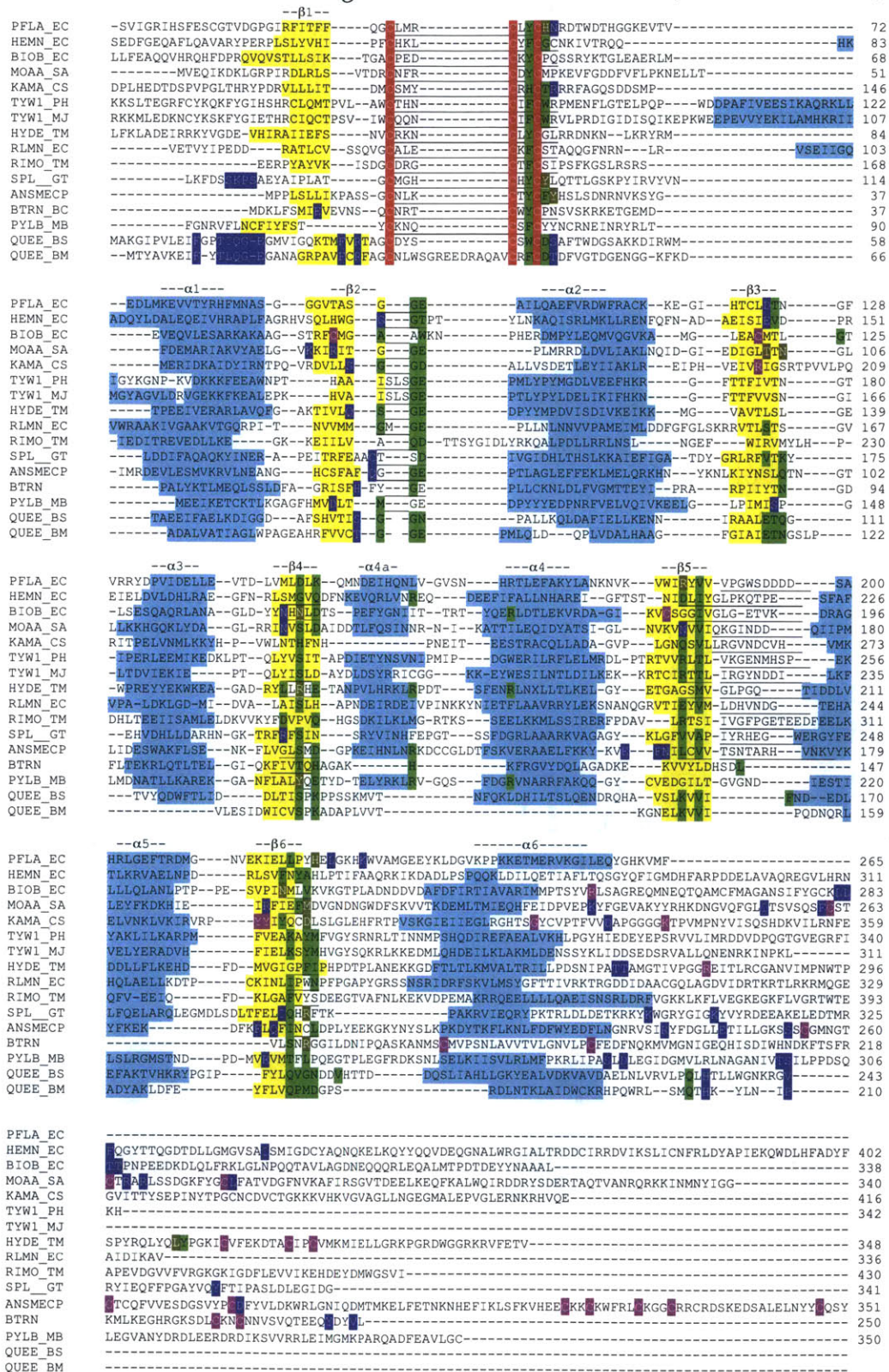
Figure V.5. Abridged AdoMet radical folds of QueE, BtrN and NrdG. (*A*) The QueE AdoMet radical fold is shown in rainbow from N-terminus (blue) to C-terminus (red). N-terminal beta strand and C-terminal helical region extensions are in grey. AdoMet and the AdoMet cluster are shown in sticks. (*B*) BtrN, colored similarly to (*A*). NrdG models generated using (*C*) BtrN and (*D*) QueE as reference models. In (*D*), AdoMet and the AdoMet cluster are modeled into the predicted structure, which has AdoMet cluster ligating cysteines and the 'GGE' (green spheres) and ribose (yellow sphere) AdoMet binding motifs in positions consistent with other AdoMet radical enzymes. The 'GXIXGXXE' and β6 motifs will most likely be housed in the β5 extension, as in BtrN (see Figure IV.6). These motifs could not be assigned from the predicted structure.
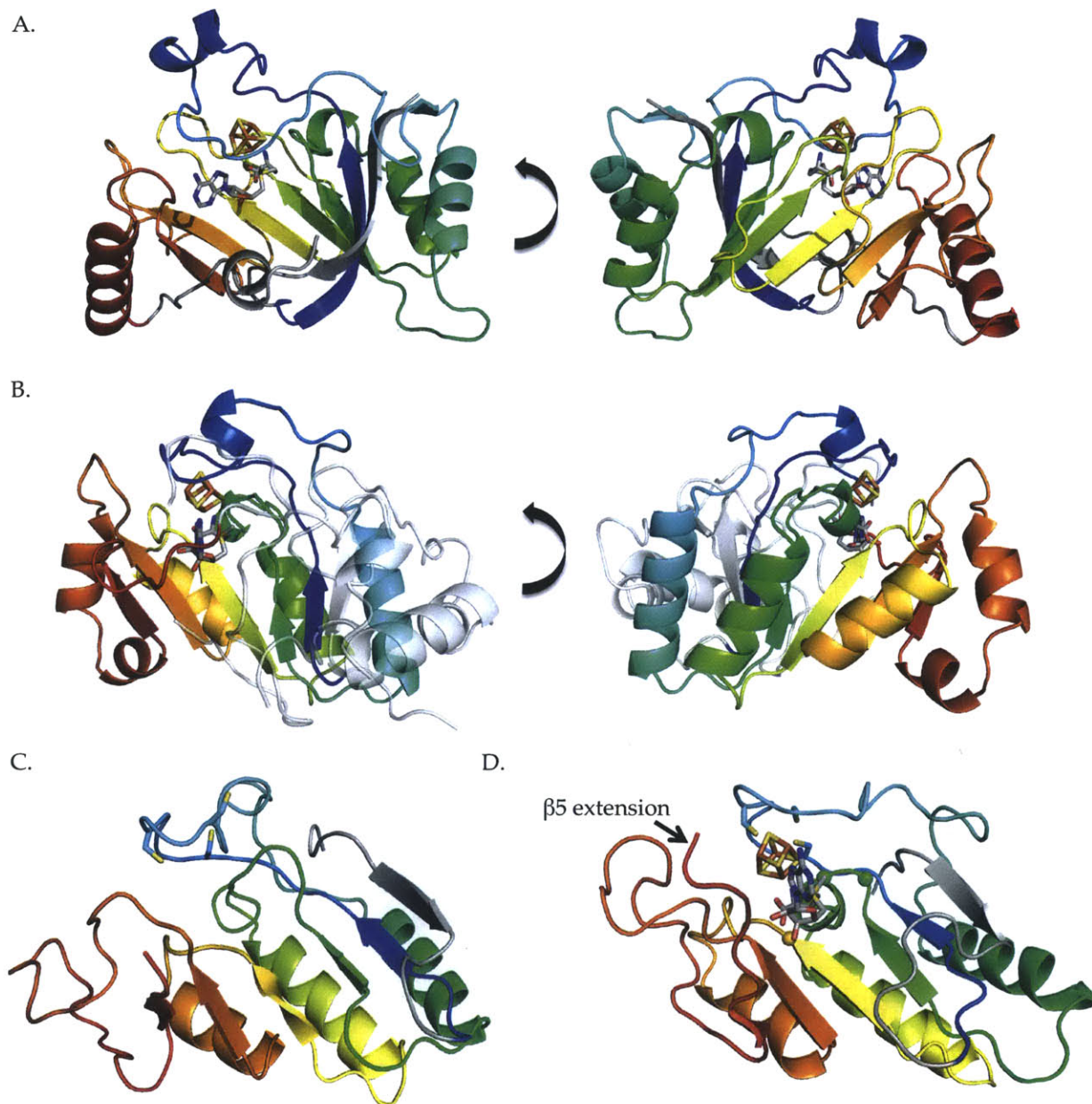
V.V REFERENCES:

1.   Lanz ND & Booker SJ (2012) Identification and function of auxiliary iron-sulfur clusters in radical SAM enzymes. *Biochim Biophys Acta* 1824(11):1196-1212.

2.   Ugulava NB, Gibney BR, & Jarrett JT (2001) Biotin synthase contains two distinct iron-sulfur cluster binding sites: chemical and spectroelectrochemical analysis of iron-sulfur cluster interconversions. *Biochemistry* 40(28):8343-8351.

3.   Ugulava NB, Sacanell CJ, & Jarrett JT (2001) Spectroscopic changes during a single turnover of biotin synthase: destruction of a [2Fe-2S] cluster accompanies sulfur insertion. *Biochemistry* 40(28):8352-8358.

4.   Tse Sum Bui B, Mattioli TA, Florentin D, Bolbach G, & Marquet A (2006) 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):3824-3834.

5.   Cicchillo RM, et al. (2004) Escherichia coli lipoyl synthase binds two distinct [4Fe-4S] clusters per polypeptide. *Biochemistry* 43(37):11770-11781.

6.   Miller JR, et al. (2000) Escherichia coli LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein. *Biochemistry* 39(49):15166-15178.

7.   Cicchillo RM & Booker SJ (2005) 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):2860-2861.

8.   Hernandez HL, et al. (2007) MiaB, a bifunctional radical-S-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe-4S] clusters. *Biochemistry* 46(17):5140-5147.

9.   Lee KH, et al. (2009) Characterization of RimO, a new member of the methylthiotransferase subclass of the radical SAM superfamily. *Biochemistry* 48(42):10162-10174.

10.   Forouhar F, et al. (2013) Two Fe-S clusters catalyze sulfur insertion by radical-SAM methylthiotransferases. *Nat Chem Biol* 9(5):333-338.

11.   Haft DH & Basu MK (2011) Biological systems discovery in silico: Radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol* 193(11):2745-2755.

12.   Fluhe L, et al. (2012) The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A. *Nat Chem Biol* 8(4):350-357.

13.   Kawulka KE, et al. (2004) Structure of subtilosin A, a cyclic antimicrobial peptide from Bacillus subtilis with unusual sulfur to alpha-carbon cross-links: formation and reduction of alpha-thio-alpha-amino acid derivatives. *Biochemistry* 43(12):3385-3395.

14.   Wecksler SR, et al. (2009) Pyrroloquinoline quinone biogenesis: demonstration that PqqE from Klebsiella pneumoniae is a radical S-adenosyl-L-methionine enzyme. *Biochemistry* 48(42):10151-10161.

15.   Grove TL, Lee KH, St Clair J, Krebs C, & Booker SJ (2008) In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry* 47(28):7523-7538.

16.   Benjdia A, et al. (2010) Anaerobic sulfatase-maturating enzyme - a mechanistic link with glycyl radical-activating enzymes? *FEBS J* 277(8):1906-1920.

17. Haft DH (2011) Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* 12:21.

18. Goldman PJ, *et al.* (2013) X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. *Proc Natl Acad Sci USA (In press)*:10.1073/pnas.1302417110.

19. Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, & Drennan CL (2004) Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* 303(5654):76-79.

20. Yokoyama K, Numakura M, Kudo F, Ohmori D, & Eguchi T (2007) Characterization and mechanistic study of a radical SAM dehydrogenase in the biosynthesis of butirosin. *J Am Chem Soc* 129(49):15147-15155.

21. Yokoyama K, Ohmori D, Kudo F, & Eguchi T (2008) Mechanistic study on the reaction of a radical SAM dehydrogenase BtrN by electron paramagnetic resonance spectroscopy. *Biochemistry* 47(34):8950-8960.

22. Grove TL, Ahlum JH, Sharma P, Krebs C, & Booker SJ (2010) A consensus mechanism for Radical SAM-dependent dehydrogenation? BtrN contains two [4Fe-4S] clusters. *Biochemistry* 49(18):3783-3785.

23. Hanzelmann P & Schindelin H (2004) Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc Natl Acad Sci USA* 101(35):12870-12875.

24. Hanzelmann P & Schindelin H (2006) 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 USA* 103(18):6829-6834.

25. Mehta AP, *et al.* (2013) Catalysis of a new ribose carbon-insertion reaction by the molybdenum cofactor biosynthetic enzyme MoaA. *Biochemistry* 52(7):1134-1136.

26. Hover BM, Loksztejn A, Ribeiro AA, & Yokoyama K (2013) Identification of a cyclic nucleotide as a cryptic intermediate in molybdenum cofactor biosynthesis. *J Am Chem Soc* 135(18):7019-7032.

27. Fluhe L, *et al.* (2013) Two [4Fe-4S] clusters containing radical SAM enzyme SkfB catalyze thioether bond formation during the maturation of the Sporulation Killing Factor. *J Am Chem Soc* 135(3):959-962.

28. Lees NS, *et al.* (2009) ENDOR spectroscopy shows that guanine N1 binds to [4Fe-4S] cluster II of the S-adenosylmethionine-dependent enzyme MoaA: Mechanistic implications. *J Am Chem Soc* 131(26):9184-9185.

29. Fugate CJ, *et al.* (2012) 9-Mercaptodethiobiotin is generated as a ligand to the [2Fe-2S]+ cluster during the reaction catalyzed by biotin synthase from Escherichia coli. *J Am Chem Soc* 134(22):9042-9045.

30. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.

31. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.

32. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, & Miller NE (2001) 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):1097-1106.

33.   Nicolet Y & Drennan CL (2004) AdoMet radical proteins--from structure to evolution--alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* 32(13):4015-4025.

34.   Roy A, Kucukural A, & Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725-738.

35.   Vey JL, *et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

36.   McCarty RM, Krebs C, & Bandarian V (2013) Spectroscopic, steady-state kinetic, and mechanistic characterization of the radical SAM enzyme QueE, which catalyzes a complex cyclization reaction in the biosynthesis of 7-deazapurines. *Biochemistry* 52(1):188-198.

37.   Smoot ME, Ono K, Ruscheinski J, Wang PL, & Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27(3):431-432.

PETER JOHN GOLDMAN

Rm. 68-688, 77 Massachusetts Avenue ◆ Cambridge, MA 02139 ◆ (206) 851-1703 ◆ pgoldman@mit.edu

EDUCATION

Massachusetts Institute of Technology, Cambridge, MA   June 2013
Ph.D. Candidate, Biological Chemistry   GPA: 4.8/5.0

Occidental College, Los Angeles, CA   May 2007
Bachelor of Arts, Chemistry, *Minors*: Physics and Mathematics   GPA: 3.8/4.0

EXPERIENCE

Research Assistant, Laboratory of Catherine Drennan, Chemistry Department, MIT   2007 - 2013
- Solved protein structures of possible drug target in collaboration with the Booker Laboratory (Penn State).
  - Designed a novel antibiotic with encouraging preliminary results and performing ongoing analysis.
- Part of a team designing and optimizing a biotechnology platform for the visualization of cellular events in collaboration with the Ting (MIT) and Baker (University of Washington) Laboratories.
- Elucidated the mechanism of an important branching point in the biosynthesis of natural products that have applications in cancer therapy and been clinically studied. A collaboration with the Walsh (Harvard Medical School) and Elliott (Boston University) Laboratories.

Undergraduate Research Fellow, Chemistry Department, Occidental College   2005-2007
- Performed temperature and pH dependent electrochemical scan rate studies with protein-film systems

PUBLICATIONS

- X-ray Structure of an AdoMet Radical Activase Reveals an Anaerobic Solution for Formylglycine Posttranslational Modification **Goldman, P.J.**, *et al. Proc. Natl. Acad. Sci. U.S.A. Accepted*   2013
- An unusual role for a mobile flavin in StaC-like indolocarbazole biosynthetic enzymes. **Goldman, P.J.**, Ryan, K.S., *et al. Chem. Biol. 19*, 855-865.   2012
- Spectroscopy and electrochemistry of cytochrome P450 BM3-surfactant film assemblies. Udit, A.K., Hagen, K.D., **Goldman, P.J.**, *et al. J. Am. Chem. Soc. 128*, 10320-10325.   2006

LEADERSHIP POSITIONS

Chemistry Graduate Student Council, MIT   2008 - 2013
- Served as President and Treasurer of student organization charged with maintaining and improving student morale and camaraderie within the Chemistry Department.
- Organized and lead multiple departmental events (first year forums and departmental town hall meetings) and social events (departmental BBQ, monthly TGIFs, and graduate student outings).
- Produced and implemented $15,000 annual budget.

Chemistry Graduate Career Panel, MIT   2012 - 2013
- Co-founder of panel series aimed at addressing the needs of graduating Chemistry PhDs with undecided career paths. Organizer of the Consulting Career Panel.

Undergraduate Research Advisor, MIT   2009 - 2013
- Develop research strategies, train, and provide guidance to multiple MIT undergraduates.

Teaching Assistant, MIT   2007 - 2008
- Taught upperclassmen biochemistry in both laboratory and classroom settings.
- Part of TA team that implemented and organized a new laboratory curriculum.

PRESENTATIONS

- Bioinorganic Chemistry Gordon Research Seminar, Ventura, CA   2013
- American Crystallographic Association Annual Meeting, Boston, MA   2012
- Bioinorganic Chemistry Gordon Research Seminar, Ventura, CA   2012
- Enzymes, Co-enzymes, and Metabolic Pathways Gordon Research Conference, Waterville Valley, NH   2011
- 240th American Chemical Society Meeting, Boston, MA   2010
- Frontiers in Metallobiochemistry, University Park, PA   2010
- Southern California Council on Undergraduate Research, Los Angeles, CA   2006

HONORS AND INTERESTS

- Phi Beta Kappa Member (2007 - present); *magna cum laude* (2007), Morse Travel Grant (2012)
- MIT Consulting Club; Intramural sports (flag football, softball, soccer, and basketball)