Two Versatile Cofactors, Flavin Adenine Dinucleotide and Non-Heme Iron, Involved in DNA Repair and Natural Product Halogenation
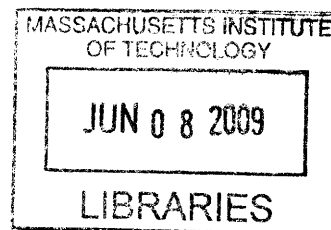
By

Cintyu Wong

B.S., Chemistry (2003)

University of Rochester

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

at the

Massachusetts Institute of Technology

May 2009

Signature of Author: _____
Department of Chemistry
May 5, 2009

Certified By: _____
Catherine L. Drennan
Professor of Chemistry and Biology, HHMI Investigator and Professor
Thesis Supervisor

Accepted By: _____
Robert W. Field
Chairman, Department 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

Professor Catherine L. Drennan

————————————————

Thesis Supervisor

Professor Leona D. Samson

————————————————

Committee Member

2

Two Versatile Cofactors, Flavin Adenine Dinucleotide an[d No]n-Heme Iron, Involved in DNA Repair and Natural Product H[alogen]ation

By
Cintyu Wong

ABSTRACT

Cofactors assist enzymes with a variety of complex chemist[ry. Two versatile cofactors,] flavin adenine dinucleotide (FAD) and non-heme iron, together [with molecular oxygen as an] oxidizing agent, perform a wide array of reactions. Hydroxylation i[n DNA repair is one example.] AidB is an adaptive response protein that is up-regulated in the p[resence of alkylating agents.] AidB contains FAD; however, the precise role of the FAD has [not been determined. AlkB,] another adaptive response protein that is up-regulated in the prese[nce of alkylating agents, is a] member of the alpha-ketoglutarate (αKG) non-heme iron-depen[dent superfamily. It uses the] cofactor non-heme iron, and the co-substrates molecular oxygen a[nd αKG, to remove alkylated] adducts on DNA bases via hydroxylation. Two well characterized [human homologues of AlkB,] hABH2 and hABH3, also belong to the αKG/Fe(II)-dependent su[perfamily, but their substrate] preference differ from that of AlkB.

Furthermore, FAD and non-heme iron, again with molec[ular oxygen as an oxidizing] agent, can perform halogenation chemistry. Flavin-dependent halo[genases perform halogenation] reactions on aromatic substrates, while non-heme iron-depe[ndent halogenases perform] halogenations reactions on unactivated aliphatic substrates. R[ebH is a flavin-dependent] halogenase that catalyzes the chlorination of free tryptophan to fo[rm chlorotryptophan in the] biosynthesis of rebeccamycin. CytC3 is an αKG and non-heme ir[on-dependent halogenase that] catalyzes the chlorination of L-2-aminobutyric acid bound to [an acylation domain in the] biosynthesis of γ,γ-dichloroaminobutyrate.

Here, we obtained the crystal structures of flavoprotein[-dependent halogenase RebH and non-heme iron-] dependent halogenase CytC3. The structural and biochemical w[ork on AidB provides new] insights into various possible functions of this still poorly und[erstood protein. The crystal] structure of CytC3 suggests two important criteria for creating an e[nzyme-bound Fe-Cl catalyst.] Additionally, we established a new purification scheme for AlkB [and its human homologues,] which has yielded protein for biochemical studies aimed at ex[ploring substrate specificity.] Finally, we have also obtained a new crystal form for *E. coli* AlkB.

To Mom, Dad and Ting

Acknowledgements

I am indebted to many professors, colleagues, friends and family m ers who have assisted me
throughout my graduate school journey. There are not enough wo o accurately describe the
amount of thanks that I owe to the many of you, but I will try her acknowledge as many as
possible.

To my advisor, Professor Cathy L. Drennan, an amazing mentor has been very patient in
fostering me as a scientist. Thank you for taking the time to guide hrough every step of the
way in the Ph.D. process, and pass on your wisdom and experienc ne. I have enjoyed being
your student and learning from you. Your enthusiasm in life and passion for science have
encouraged me to never give up. I thank you for everything that yo 'e done to prepare me for
my career. A special thanks to you for working very hard with me e past months to help me
finish my thesis defense.

To my thesis committee, Professor John M. Essigmann and Leo . Samson, who are very
knowledgeable about DNA repair and cancer research. I thank yo all the time that each of
you has spent on me, and for your thoughtful discussions abou projects. I have always
walked away from your offices with a greater knowledge an riosity about science. I
appreciate both of your help and care during my job search for the r tep in my career.

To the current and former Drennan lab members who have read ai ovided comments on my
thesis, including Christine Phillips, Yan Kung, Leah Blasiak, Bec Bjork, Mishtu Dey, Peter
Goldman and Danny Yun, thank you for volunteering your time. uld not have completed
without your help.

To all of the Drennan lab members who have made this place a fui king environment, thank
you for your friendship and willingness to share. I learned sc ling from everyone. To
Christine Phillips and Yan Kung, thank you for all the help that I received from you both,
and I enjoyed the late nights at the synchrotron together. You g vill be doctors soon! To
Danny Yun, thank you for setting up and teaching me all the cc ter programs. To Becky
Bjork, thank you for helping me fundraise for the Avon Walk for B Cancer and cheering me
on. To Nosomi Ando and Mishtu Dey, thank you for sharing your rience and advice on the
next step after graduate school. To Christina Stock, you are cou us to work on the RNR
complex, and thank you for sharing your general knowledge of ie many topics that you
know. To Peter Goldman, thank you for organizing the bowlin o and I am sorry that I
couldn't make it.

To my great collaborators, Jim Delaney, Deyu Li and Mike Hamil io have generously spent
numerous hours discussing AlkB and AidB with me, and taught n ur expertise. I appreciate
your help on developing the AidB story when I was under a strict d ne.

To my fellow Chemistry classmates, thanks for memories of going through classes and orals together. Special thanks to Jiejin Chen, Jialu Yeh, and Wan-Chen Lin, for all the good times that we had together the past few years, and our long talks over dinner. To Jiejin, your independence and your ability to be strong always amaze me. You taught me not to take anything for granted. To Jialu, I value your optimism and courage. You have been very generous to me in a way that I cannot repay. Thank you for always keeping me in your prayers. To Wan-Chen, I am in awe of your passion for science and teaching; you will be a great teacher. Thank you for giving me hope when my experiments failed, and for feeding me cake when we ate lunch together.

To my parents, who have always believed in me and encouraged me to do what I love. Thank you for many years of patience and sleepless nights to raise me and educate me. You two gave up a comfortable life in China and moved to the USA for me, and worked very hard to provide the best education and living environment for me. To my mother, who is also my best friend, you are my role model; you taught me to be strong and be optimistic in life. You gave me reasons to come to graduate school and you have been there for me and gave me courage to get through every hurdle. To my dad, thank you for your unconditional love and unselfish patience with me. You are always proud of me and believe in my abilities even when I doubt myself.

To my wonderful boyfriend Ting, you have provided companionship, friendship, love, and support at a time when I needed them most. Thank you for your advice, patience, and humor to get me through the stressful times, and I am grateful for your not walking out on me when I vented my frustrations, complaints, and anger on you. I have enjoyed sharing successful and happy moments of my life with you, as well as doing mundane everyday tasks with you. I hope this is not the end of our journey together, but the beginning.

TABLE OF CONTENTS   PAGE

Chapter IV: Crystallographic Analysis of an Open Active Site Conformation of Non-Heme Iron Halogenase CytC3

LIST OF FIGURES   PAGE

11

Chapter I

Introduction to the Use of Flavin and Non-heme Iron as Cofactors in DNA Repair and Halogenation Chemistry

I.A. Summary

Cofactors assist enzymes with a variety of complex chemistries. Two versatile cofactors, flavin adenine dinucleotide (FAD) and non-heme iron, together with molecular oxygen as an oxidizing agent, perform a wide variety of reactions, ranging from DNA repair to halogenation. The respective roles of FAD and non-heme iron in enzymes that perform hydroxylation in DNA repair and halogenations reactions are discussed in this chapter. A few enzymes that use these cofactors are discussed in detail: AidB is a flavoprotein that is up-regulated in response to DNA damaging agents, but whose function is unknown, AlkB and its human homologues use non-heme iron to perform hydroxylation reactions in DNA repair, and CytC3 uses non-heme iron to perform halogenation reactions.

I.B. Introduction

Enzymes, composed of carbon, nitrogen, oxygen and sulfur, can perform chemical reactions with great efficiency. However, for some complex chemistry, an enzyme may require the help of cofactors. Cofactors are small molecules or metals that are bound to the protein and expand the potential biological activity. This thesis will focus on the use of flavin adenine dinucleotide (FAD) and non-heme iron as cofactors in DNA repair and halogenation chemistry.

DNA repair plays an essential role in maintaining genomic integrity. DNA damage arises from both endogenous and exogenous sources, such as methylation by the natural coenzyme *S*-adenosylmethionine and methyl methanesulfonate in the environment, respectively. The damaged DNA must be repaired efficiently in order to prevent it from causing cellular mutagenicity and cytotoxicity. All organisms have multiple enzymes to cover the repair of 13 methylated DNA lesions;[1] different DNA repair enzymes have different lesions as substrates, and they exploit a variety of mechanisms. In *E. coli*, AidB is a flavin-binding adaptive response protein whose expression is up-regulated in the presence of DNA alkylating agents.[2] Another DNA repair enzyme that is part of the same *E. coli* pathway is the AlkB protein. In humans, there are 8 different AlkB homologues (hABH1-8). AlkB and its human homologues are part of the alpha-ketoglutarate (αKG) non-heme iron-dependent superfamily.[3, 4] Enzymes in this superfamily catalyze a range of diverse reactions, including epimerization, cyclization, epoxidation, halogenation, and hydroxylation, among others. Specifically, AlkB and its homologues belong to the hydroxylase subclass of the αKG/Fe(II)-dependent superfamily. The putative repair proteins AidB, repair protein AlkB, and two human homologues of AlkB (hABH2 and hABH3) will be discussed in this thesis.

Halogenation chemistry is important for the synthesis of many natural products that have therapeutic properties. Over 4000 natural products are halogenated, where chlorination is the most common, followed by bromination.[5] Examples of chlorinated natural products include salinisporamide A, a potent anti-cancer agent, and syringomycin, an anti-fungal compound.[6, 7] The incorporation of a halogen into natural products can be important for tailoring a compound's activity. For example, the addition of a single chlorine atom to the syringomycin scaffold increases its anti-fungal properties four-fold.[8] There are four types of cofactor-mediated halogenating enzymes: heme iron haloperoxidases, vanadium-dependent haloperoxidases, flavin-dependent halogenases, and non-heme iron halogenases.[9] The latter two types of halogenating enzymes, the flavin-dependent and the non-heme iron halogenases, use molecular oxygen as an oxidizing agent in conjunction with their reactive cofactors. CytC3 is a recently discovered chlorinating enzyme that belongs to the halogenase subclass of the αKG non-heme iron-dependent superfamily.[10] The second part of the thesis will examine the subtle differences that are important for creating an enzyme active site that binds halogen to perform the halogenation reaction.

In this introductory chapter, the following topics are discussed: (1) an introduction to flavin chemistry in biological systems, (2) reactivity and structural features of the αKG/Fe(II)-dependent enzymes, (3) the adaptive response to alkylating agents in *E. coli* for DNA repair, (4) an adaptive response flavoprotein—AidB, (5) αKG/Fe(II)-dependent DNA repair enzymes—AlkB and human homologues of AlkB, (6) the role of flavin and non-heme iron in halogenations, (7) a flavin-dependent halogenase—RebH, and (8) an αKG/Fe(II)-dependent halogenase—CytC3.

I.C. Introduction to Flavin Chemistry in Biological Systems

Flavin adenine dinucleotide (FAD) is a versatile redox cofactor used for electron transport processes and redox reactions. FAD is usually bound to proteins and cycles between oxidized and reduced states during enzyme catalysis. The redox potential of FAD ranges greatly depending on the protein environment; this wide range of redox potential allows FAD to participate in one-electron processes such as radical reactions or two-electron processes such as oxidation reactions.[11] In addition, reduced flavin ($FADH_2$) can be oxidized by molecular oxygen, which allows FAD-dependent enzymes to participate in hydroxylation reactions. A few of the diverse roles of FAD in biological systems are oxidation, dehydrogenation, hydroxylation, and halogenation.

Oxidation of various chemical groups often depends on a redox-active role for FAD. For example, oxidation of hemiacetals to lactones is catalyzed by glucose oxidase, oxidation of alcohols to aldehydes is catalyzed by glycolate oxidase, and oxidation of amines to imines is catalyzed by amino acid oxidase.[12] These oxidases use FAD as a cofactor to oxidize the substrate. The oxidases regenerate FAD through rapid reaction of $FADH_2$ with molecular oxygen to generate a C4a-hydroperoxy FAD intermediate. The C4a-hydroperoxy FAD breaks down to hydrogen peroxide and oxidized FAD in the protein. Another type of oxidation is the dehydrogenation reaction, or oxidation of carboxylic acids to α,β-unsaturated carbonyl compounds. Many acyl-CoA dehydrogenases catalyze this type of reaction using FAD. This type of oxidation does not require molecular oxygen. A Glu residue from the protein initiates the α-carbon deprotonation on the substrate, followed by a hydride shift from the β-carbon of the substrate to FAD, forming reduced FADH⁻ (Scheme I.1).[13] The FADH⁻ is reoxidized by another FAD-dependent protein called the electron transfer flavoprotein (ETF).

In both hydroxylation and halogenation reactions, fully reduced FAD reacts with molecular oxygen to generate the C4a-hydroperoxy FAD intermediate, just as the oxidative half-reaction in an oxidase. For hydroxylation, the C4a-hydroperoxy FAD intermediate does not decompose into hydrogen peroxide and oxidized FAD as in oxidase reactions; instead the intermediate transfers a hydroxyl group to the substrate (Scheme I.2). The resulting C4a-hydroxy FAD returns to oxidized FAD upon dehydration. In halogenations, the C4a-hydroperoxy FAD intermediate oxidizes a halide ion to produce hypohalous acid (Scheme I.2). The hypohalite then further transfers the halide to the substrate.[9]

I.D. αKG/Fe(II)-Dependent Superfamily

The αKG/Fe(II)-dependent superfamily appears to be one of the most versatile families of enzymes. These enzymes require molecular oxygen as a cosubstrate to oxidatively decarboxylate αKG, which is then coupled to an oxidation reaction of the substrate. Carbon dioxide and succinate are generated as products. Reactions catalyzed by some of the family members include desaturation, ring expansion, ring formation, epoxidation, hydroxylation, halogenation, and other oxidative reactions.[14] For example, clavaminate synthase (CAS) catalyzes three steps: the hydroxylation of deoxyguanidinoproclavaminic acid, the formation of cyclic rings, and desaturation reactions of proclavaminic acid in the synthesis of clavulanic acid (Scheme I.3.a);[15] deacetoxycephalosporin C synthase (DAOCS) performs both ring expansion and hydroxylation in cephalosporin synthesis starting with penicillin N (Scheme I.3.b);[16] taurine hydroxylase (TauD) hydroxylates taurine (Scheme I.3.c);[17] and CytC3 chlorinates aminobutyric acid (Scheme I.3.d).[10] The substrates for these enzymes range from large biomolecules, such as proteins and DNA/RNA, to small molecules, such as lipids and metabolites. As versatile as the

functions of these enzymes are, they all use αKG and iron as cofactors, and they share a conserved structural core.

All αKG/Fe(II)-dependent enzymes share a common fold, referred to as a cupin fold, comprising anti-parallel β-strands in a "jelly-roll" motif (Figure I.1).[18] Each enzyme differs in the additional α-helices and protein loops surrounding this common core. The iron cofactor is located in the center of the "jelly-roll" and is coordinated by a conserved 2-His-1-carboxylate (His-X-Asp/Glu-Xn-His motif) triad.[19] These three conserved residues ligate the iron center on one face of the octahedral site, and the other three sites are ligated by water when the enzyme is in the resting state (Figure I.2). There is one exception to this conserved 2-His-1-carboxylate triad, and that is in the halogenase subclass, where the carboxylate ligand is absent. Alanine replaces the aspartate/glutamate of the 2-His-1-carboxylate motif in the halogenases, opening an additional coordination site on the iron.[20] The active site of halogenases must be able to bind a halogen for the reaction to occur; the small size of the alanine substitution and the non-coordinating nature of the side chain allow room for a halogen to bind directly to iron in the position that is typically occupied by the side chain carboxylate. The αKG cofactor binds to the iron in a bidentate fashion through the C1 carboxylate and the keto group on C2 (Figure I.2 for numbering nomenclature). Bidentate binding of αKG to iron is strictly conserved in the αKG/Fe(II)-dependent enzyme family.[21]

A large number of spectroscopic, biochemical, and structural studies have been done on the taurine hydroxylase enzyme TauD, and a general mechanism for this class of enzymes has been proposed from these results.[22-24] The resting state of the enzyme has three protein ligands and three water molecules coordinating the iron (Figure I.2). αKG addition displaces two water molecules when it binds in a bidentate fashion to the iron center.[25] This replacement of neutral

18

water molecules with anionic ligands decreases the Fe(III)/Fe(II) reduction/oxidation potential, leaving the iron center more susceptible to oxidation.[19] Binding of the substrate causes the displacement of the remaining water molecule, thereby creating a five-coordinate iron center, which primes the site for binding molecular oxygen. Molecular oxygen does not bind until the substrate and/or cofactors are bound, which serves to prevent self-inactivation in this family of enzymes.[26] In a typical reaction, the uncoordinated oxygen atom of dioxygen attacks the C2 carbonyl of αKG, which promotes cleavage of the O-O bond, thus concomitantly forming the ferryl-oxo species (Fe(IV)=O). The ferryl-oxo species has been observed in a number of enzymes in this family, including TauD and CytC3, and is believed to be responsible for hydrogen abstraction from the substrate.[22-24, 27, 28] The substrate radical recombines with a hydroxyl radical to give the hydroxylated product. Product release and rehydration of the iron center return the enzyme active site to the resting state, ready for another round of catalysis.

I.E. Adaptive Response for DNA Repair

DNA methylation by either S_N1 or S_N2 alkylating agents can occur at 13 different sites: 12 methylation sites are on the DNA bases, including all exocyclic oxygens and most cyclic nitrogens; one methylation site occurs on the DNA phosphate backbone (Figure I.3).[1] The adaptive response is triggered when low levels of alkylating (mainly methylating) agents are present, thereby providing resistance to cell death caused by higher levels of such alkylating agents.[29, 30] The adaptive response is a cellular process that protects against cell death by utilizing various cellular machineries that repair methylation damage. There are three enzyme families involved in repairing methylated DNA lesions, and each is up-regulated during the adaptive response: methyltransferases, DNA glycosylases, and DNA dioxygenases.

The Ada protein is a DNA repair methyltransferase in *E. coli* that also acts as a transcription factor that regulates the adaptive response.[31] Methyltransferases repair methylation on the major groove exocyclic oxygens, providing resistance to S_N1 alkylating agents. The Ada protein is a bifunctional protein that is composed of two domains: an N-terminal domain that demethylates the methylphosphotriester in DNA and a C-terminal domain that demethylates O^6-methylguanine (O^6-mG) and O^4-methylthymine (O^4-mT).[32] Ada demethylates DNA lesions by direct transfer of the methyl group from the damaged base to the thiol group of a Cys residue in the active site. Upon methylation of the active site Cys in the N-terminal domain of Ada, the protein behaves as a transcription factor, and its affinity to the *ada-alkB*, *alkA*, and *aidB* operons greatly increases, in turn, up-regulating the expression of Ada, AlkB, AlkA, and AidB proteins (Figure I.4).[1, 31]

AlkA is a DNA glycosylase that is induced as part of the adaptive response.[33] DNA glycosylases generally repair lesions that form destabilized glycosylic bonds. These destabilized bonds are caused by both S_N1 and S_N2 alkylating agents. Unlike other DNA repair enzymes that reverse methylation directly, DNA glycosylases work by hydrolyzing the glycosylic bond to yield an abasic site.[1] Therefore, DNA glycosylases require the help of other proteins, such as endonuclease, polymerase, and ligase, to complete the repair of damaged lesions. After the hydrolysis of the methylated base, endonucleases cleave the DNA backbone.[34] A normal base is restored by a DNA polymerase based on the sequence of the complementary strand, and a DNA ligase joins the new base with the rest of the DNA strand.[1]

Another enzyme that is up-regulated in the presence of alkylating agents is the AidB protein.[2] Although AidB is known to be part of the adaptive response, its exact function has not been discovered. From cell survival experiments, there is evidence for AidB providing resistance

to S_N1 alkylating agents;[2] and it is hypothesized that AidB can detoxify alkylating chemicals that cause DNA damage.[35, 36] AidB is a flavin-containing protein[37], but it is unclear if the flavin is structural or catalytic.

AlkB is a DNA dioxygenase that removes methyl groups from the nitrogen of the heterocyclic bases involved in base pairing, providing resistance to S_N2 alkylating agents.[31] AlkB is a member of the αKG/Fe(II)-dependent dioxygenase family.[38] It utilizes αKG and molecular oxygen as a co-substrates, and non-heme iron as cofactor, to produce succinate and formaldehyde. The decarboxylation of αKG is coupled to the hydroxylation of the methyl group on the DNA lesion.[3, 4] The resulting hydroxylated methyl group is a much better leaving group, which is thought to decompose to formaldehyde and the unmethylated base (Scheme I.4).

I.E.1. An Adaptive Response Flavoprotein—AidB

AidB is an *E. coli* adaptive response protein.[2] This protein was discovered several decades ago; however, the true function of AidB is still a mystery. The induction of AidB expression has been shown to be dependent on the adaptive response. AidB is overexpressed in the presence of N-methyl-N′-nitro-nitrosoguanidine (MNNG), an S_N1 methylating agent.[2, 36] The fact that AidB responds to MNNG specifically leads to the hypothesis that AidB plays an active role in destroying the alkylating agent MNNG, providing resistance to it.[2, 35, 39] This hypothesis is supported by knock out studies by Volkert *et al.*; where AidB knock out cells are more sensitive to MNNG in a cell survival assay, which suggests that the presence of AidB provides protection against cell death caused by MNNG.[2] However, in the thesis work of Lauren Frick from Professor John Essigmann's laboratory, this result was not reproducible.[40]

Independent of the adaptive response, AidB expression is also known to be induced when cells are in an anaerobic environment or in the presence of acetate at a slightly acidic pH (pH 6.5).[41-43] The significance of AidB induction under these conditions is unclear.

Sequence alignment predicts that AidB belongs to the acyl CoA dehydrogenase superfamily (Figure I.5). In particular, AidB is functionally homologous to isovaleryl CoA dehydrogenase (IVD), which is an enzyme involved in one of the main steps of leucine metabolism. IVD catalyzes the α,β-oxidation of 3-methylbutanoyl CoA to 3-methylcrotonyl CoA.[44] IVD is a homotetramer in solution and it contains one FAD molecule per monomer. Two FADs bind at the interface of an IVD dimer (one per monomer), in the extended conformation. A glutamate residue (Glu254) is the active site base that initiates proton abstraction from the substrate (Scheme I.1).[44]

Biochemical studies show that AidB binds FAD.[37] However, whether the role of FAD in AidB is structural or catalytic remains unclear because no true function for the enzyme has been identified. Hausinger *et al.* suggested a possible reaction for AidB based on the following: (1) AidB's sequence similarity to IVD, suggesting AidB performs some kind of dehydrogenation reaction; (2) AidB's association with the adaptive response; and (3) AidB's ability to bind DNA.[37] Together, (2) and (3) suggest AidB performs some kind of DNA repair reaction. Putting these clues together, a hypothetical function of AidB is the use of dehydrogenase activity to repair alkylated DNA (Scheme I.5).[37] In this proposed reaction, AidB forms a double bond between the methyl group and the normal base of a methylated DNA lesion by dehydrogenase activity and the Schiff base can be hydrolyzed by water to release as formaldehyde. Although this proposed reaction for AidB is attractive, there is no experimental evidence for it.

A recent structural study suggests that AidB is a homotetramer, with two FADs bound at the interface of two monomers.[45] Based on crystal packing and biochemical study, Eichman *et al.* proposed that double-stranded DNA (dsDNA) binds to AidB on two sides of the homotetramer, and four tetramer units can arrange themselves so that they seclude dsDNA in a 25 Å pore (Figure I.6).[45] This model of AidB protecting dsDNA from alkylating agent seems unlikely for a number of reasons: (1) it requires a large concentration of AidB in the cell, (2) dsDNA is usually present as a supercoil and thus is unable to bind to AidB as suggested in the model, and (3) the dsDNA must be desolvated before being secluded by AidB, and such desolvation would be very energetically unfavorable. In Chapter II of this thesis, the crystal structure of AidB and supporting biochemical work will be presented, and various proposals for the function of AidB will be discussed.

I.E.2. Adaptive Response αKG/Fe(II)-Dependent DNA Repair Enzymes—AlkB and Its Human Homologues

AlkB is a DNA repair enzyme that belongs to the αKG/Fe(II)-dependent dioxygenase family.[1] AlkB directly reverses DNA damage via oxidative demethylation.[1] Specifically, AlkB repairs methylated DNA adducts on the nitrogen of the heterocyclic bases involved in pairing.[1] AlkB couples the decarboxylation of αKG to the hydroxylation of the methyl group on the DNA lesion (Scheme I.4). The subsequent hydroxylated methyl group is unstable, and is thought to decompose spontaneously to formaldehyde and the normal base.

Sequence searches found eight human homologues (hABH1-8) of AlkB, also putative members of the αKG/Fe(II) dioxygenase family.[46] hABH2-3 are the most studied, and their activities are most similar to AlkB. For many years, hABH1 has been shown to provide

resistance to methyl methanesulfonate (MMS), an S_N2 alkylating agent, *in vivo*; however, a specific substrate has not been identified. Recently, a study shows that hABH1 demethylates 3-methylcytosine (3mC) on a single-stranded DNA (ssDNA) as well as on RNA.[47] Few studies have been done on the other human homologues (hABH4-8), and their substrates have not been identified. For this reason, the focus of Chapter III of this thesis will be on AlkB, hABH2, and hABH3.

Although both the *E. coli* DNA repair protein, AlkB, and the human homologues, hABH2 and hABH3, repair the same pool of methylated DNA lesions, they each have somewhat different substrate preferences. AlkB repairs a large collection of lesions including methylated lesions, such as 1-methyladenine (1mA), 3mC, 3-methylthymine (3mT), and 1-methylguanine (1mG), as well as alkylated lesions, such as $1,N^6$-ethenoadenine (ϵA) and $1,N^6$-ethanoadenine (EA).[3, 4, 48-51] AlkB repairs these lesions on both ssDNA and RNA.[3, 4] The two human homologues have narrower substrate specificity compared to AlkB. hABH2 repairs 1mA on dsDNA, but it does not repair RNA.[52, 53] A recent study reports that hABH2 also repairs ϵA on dsDNA.[54] hABH3 repairs 1mA and 3mC on ssDNA and RNA (Table I.1).[52, 53]

The crystal structures of AlkB,[55] hABH2,[56] and hABH3[57] were recently solved. These structures, some with DNA substrates, can explain the differences in the preference for ssDNA versus dsDNA. These structures contain a core fold similar to the common "jelly roll" motif found in the αKG/Fe(II)-dependent superfamily.[55-57] The active site iron is coordinated by the conserved 2-His-1-carboxylate triad. The crystal structure of AlkB shows a region (residues 45-90) that is involved in interacting with the substrate and covering the active site.[55] Amide hydrogen exchange studies have shown this region of the structure is flexible, which may explain the broad spectrum of substrates AlkB repairs.[55]

He *et al.* reported the structures of AlkB and hABH2 cross-linked to 1mA-dsDNA.[56] The 1mA was modified with a disulfide cross-linking group. A Cys mutation for crosslinking purposes was introduced at a residue near the active site, D135C for AlkB and E175C or G129C for hABH2. Several other surface Cys to Ser mutations were introduced to eliminate nonspecific cross-linking. This study shows that AlkB and hABH2 have different base flipping mechanisms under artificial constraints linking the protein and the substrate together. In AlkB, the DNA bases on either side of the lesion are squeezed together to maintain base stacking after the lesion is flipped out into the enzyme's active site, and there is no protein-assisted stabilization of the base stacking in the DNA helix.[56] On the other hand, hABH2 inserts F102 into the empty space created by base flipping to stabilize the DNA helix.[56] According to structural alignment of AlkB and hABH2, the F102 in hABH2 is missing in AlkB. Currently, there is no structure of hABH3 with DNA bound. Some biochemical work on AlkB, hABH2 and hABH3, and structural work on AlkB will be discussed in detail in Chapter III.

I.F. The Role of Flavin and Non-heme Iron in Halogenation

As previously alluded to, many halogenated natural products have therapeutic properties. Nature uses four different cofactors for halogenations reactions: heme iron, vanadium, flavin and non-heme iron.[9] Two of these cofactors, flavin and non-heme iron, assist halogenation reactions in combination with molecular oxygen. Halogenases that use flavin or non-heme iron cofactors are mostly found in polyketide synthetase (PKS) and nonribosomal peptide synthetase (NRPS) assemblies. The first flavin-dependent halogenase discovered was PrnA, which catalyzes the chlorination of free tryptophan to form 7-chlorotryptophan in the biosynthesis of pyrrolnitrin (Scheme I.6.a).[58] Similar enzymes that also chlorinate free tryptophan were later discovered:

RebH catalyzes the same reaction as PrnA, but in the biosynthesis of rebeccamycin;[59] PyrH catalyzes the chlorination of free tryptophan at the 5-position in the biosynthesis of pyrroindomycin;[60] and ThaI chlorinates free tryptophan at the 6-position in the biosynthesis of thienodolin.[9] Amino acids attached to the thiolation (T) domain of the NRPS are also known to be halogenated by flavin-dependent halogenases. For example, halogenase PltA catalyzes the di-chlorination of the pyrrole ring of the oxidized proline at the C4 and C5 position when attached to a T-domain in the biosynthesis of pyoluteorin (Scheme I.6.b).[61, 62] All of these flavin-dependent halogenases catalyze halogenation of aromatic rings.

These flavin-dependent halogenating enzymes use molecular oxygen to oxidize the flavin to a C4a-hydroperoxy flavin intermediate.[63] This intermediate then oxidizes a halide ion to produce hypohalous acid.[63] The hypohalite is believed to be attached to a protein side chain, and thus restricts the regiospecificity of the reactive intermediate that halogenates the substrate.[64]

While hypohalous acid is ideal for halogenating aromatic moieties, such as tryptophan and pyrrole, the oxidizing power of hypohalous acid is not strong enough to halogenate unactivated aliphatic carbons; therefore, the flavin-dependent halogenase is not expected to work on such substrates.[9] For halogenations of unactivated aliphatic carbons, nature uses the non-heme iron halogenases, found in the NRPS assembly line. Both SyrB2 and CytC3 are examples of non-heme iron halogenases. SyrB2 catalyzes the chlorination of L-threonine bound to a thiolation domain in the biosynthesis of syringomycin E (Scheme I.7),[65] and CytC3 catalyzes the chlorination of L-2-aminobutyric acid (Aba), also attached to a thiolation domain (Scheme I.3.d).[10] Biochemical studies found that the non-heme iron halogenase requires αKG, Fe(II), chloride, and molecular oxygen for catalysis.[65] The non-heme iron halogenases are a subclass of the αKG/Fe(II)-dependent superfamily.

As the previous two sections mention, AlkB also belongs to the αKG/Fe(II)-depedent superfamily, in the most common hydroxylase subclass. A major difference between hydroxylases and halogenases is that the hydroxylases coordinate the iron with the typical 2-His-1-carboxylate (Asp/Glu) motif, whereas the halogenases coordinate the iron with only two histidines.[20] Alanine takes the place of the aspartate/glutamate of the 2-His-1-carboxylate (Asp/Glu) motif in the halogenases; the small size of the alanine substitution and the non-coordinating nature of the side chain allow room for chloride to bind directly to iron in the position that is typically occupied by the side chain carboxylate.[66] In addition to the protein ligands, the remaining coordination sites of iron are taken by chloride, αKG and molecular oxygen.

I.F.1. Flavin-Dependent Halogenase—RebH

Rebeccamycin is a chlorinated antitumor agent.[67] One of the early steps in the biosynthesis of rebeccamycin is the chlorination of L-Trp to form 7-chlorotryptophan, which is catalyzed by FAD-dependent halogenase RebH.[59] The chlorination reaction requires $FADH_2$, molecular oxygen, and chloride ion. Another protein, RebF, catalyzes the NADH-dependent reduction of FAD to regenerate the $FADH_2$ which then binds to RebH.[59]

As the previous section discussed, $FADH_2$ reacts with molecular oxygen to generate a C4a-hydroperoxy FAD intermediate. This oxidative half reaction of $FADH_2$ does not require the substrate of RebH—tryptophan—nor does the rate of the C4a-hydroperoxy FAD intermediate formation change in the presence of tryptophan, suggesting that the flavin intermediate does not directly react with tryptophan.[63] Instead, the flavin intermediate must react with chloride ion to

form hypochlorous acid (HOCl) and C4a-hydroxyl FAD.[63] The dehydration of the C4a-hydroxy FAD generates FAD.

HOCl is a powerful enough oxidant to chlorinate tryptophan.[64] However, since HOCl is a small molecule that can diffuse through the proteins easily, it is difficult to believe that HOCl can carry out chlorination with specific regiochemistry only at the 7-position of tryptophan. Yeh *et al.* have shown a long-lived, up to 63 h at 4°C, chlorinating intermediate that is covalently attached to RebH.[64] The structure of RebH suggests that a lysine (Lys79) could react with HOCl to form a protein bound lysine chloramine, which then further transfers the chloride to the substrate with regiospecificity.[64] Mutation of Lys79 results in a loss of the covalently bound chlorinating intermediate and inactivates the enzyme. It is thus believed that halogenation through a protein bound chloramine is less reactive than the free hypochlorous acid but provides more regioselectivity.

I.F.2. αKG/Fe(II)-Dependent Halogenase—CytC3

There are many therapeutic natural products containing a halogen atom in their final structure, and the biological activity that this halogen atom contributes can be significant.[5] Another role for halogenations is in the biosynthesis of natural products with cyclopropane ring structures; these reactions involve so-called "cryptic halogenations" where a halide is used to functionalize a non-reactive group and is subsequently displaced by a substrate-derived enolate to generate the three-membered ring.[68] Since chlorination by chemical synthesis can be environmentally unfriendly and technically challenging, there is interest in understanding enzymatic halogenation chemistry for the production of various natural products.

28

The biosynthesis of 1-aminocyclopropane-1-carboxylate (ACC) from L-2-aminobutyric acid (Aba) uses non-ribosomal peptide synthetase (NRPS) modules involving cryptic chlorination, reminiscent of that for coronamic acid.[10] The gene cluster for ACC biosynthesis contains four open reading frames, CytC1–C4.[10] CytC1 is a free-standing adenylation domain that loads the nonproteinogenic amino acid Aba onto the phosphopantetheine arm of a thiolation domain, CytC2 (Scheme I.8). While the amino acid is tethered to CytC2, CytC3 then catalyzes tandem chlorinations of the γ-methyl of Aba. Presumably, the tethered chlorinated Aba would form ACC by intramolecular displacement of the chloride. This step has been shown to be catalyzed by CmaC, an enzyme that facilitates the cyclization to form a cyclopropyl ring on a tethered substrate.[68] Lastly, CytC4 is the thioesterase that hydrolytically releases chlorinated aminobutyrate from the phosphopantetheine arm of CytC2. ACC is postulated to be incorporated into an apoptosis inducer, Cytotrienin A, in *Streptomyces* soil bacteria.[69]

While Aba is the natural substrate for the CytC biosynthetic pathway, other amino acids can also be adenylated by CytC1. By measuring the reversible amino acid-dependent exchange of $^{32}PP_i$ into ATP, Masashi and coworkers determined that five other amino acids, L-valine, ACC, D-Aba, D-valine, and 4-Cl-L-Aba (the product of CytC3) are adenylated better than the natural substrate, L-Aba (Table I.2).[10] This data suggests that CytC1 is a promiscuous adenylation domain, which is surprising because most other adenylation domains are relatively substrate-specific. For example, the adenylation on the biosynthetic pathway of coronamic acid, CmaA, has high specificity for L-Ile.[68]

Many spectroscopic studies have been done on CytC3. These include freeze-quench Mössbauer spectroscopy to characterize the presence of an iron(IV)-oxo intermediate, stopped-flow absorption to study the kinetics of CytC3 reaction, and circular dichroism (CD) and

magnetic circular dichroism (MCD) to reveal the structural effects of the perturbation of carboxylate (Asp/Glu) to Ala mutation in the active site.[27, 28, 70] The stopped-flow absorption and the freeze-quench Mössbauer spectroscopies identify an intermediate in the catalytic cycle that is comprised of two high-spin Fe(IV) states in rapid equilibrium. Similar high-spin Fe(IV) intermediate states are also observed in TauD, which is the model system for non-heme iron hydroxylase studies. One of the intermediate species is the high-spin Fe(IV) oxo species that is responsible for hydrogen abstraction.[27] The second intermediate species in rapid equilibrium is the Fe(II)-containing enzyme-product complex.[27]

The CD study shows that apo CytC3/Cl⁻ does not bind iron; however, 0.9 equivalents of iron binding can be detected in the presence of αKG.[70] This result is in agreement with metal analysis of SyrB2 reconstituted both without and with αKG, which contained 0.3-0.4 and 0.85-0.95 mole Fe(II)/mole subunit, respectively.[10, 65] The MCD study shows that the αKG/Fe(II)/Cl⁻ complex of CytC3 contains a weakly coordinated water ligand, suggesting the complex is ready to bind oxygen.[70] The weakly coordinated water ligand is also observed in TauD; however, a strongly coordinated water ligand is observed in clavaminate synthase (CAS). These spectroscopic studies provide a deeper understanding of the local environment of the CytC3 active site. The crystal structure of CytC3, discussed in detail in Chapter IV, will give insights into the important criteria for creating an enzyme-bound Fe-Cl catalyst.

I.J. References

1. Sedgwick, B., Repairing DNA-methylation damage. *Nature Reviews Molecular Cell Biology* **2004,** 5, (2), 148-157.

2. Volkert, M. R.; Nguyen, D. C., Induction of Specific Escherichia-Coli Genes by Sublethal Treatments with Alkylating-Agents. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1984,** 81, (13), 4110-4114.

3. Falnes, P. O.; Johansen, R. F.; Seeberg, E., AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **2002,** 419, (6903), 178-182.

4. Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B., Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002,** 419, (6903), 174-178.

5. Walsh, C., Molecular mechanisms that confer antibacterial drug resistance. *Nature* **2000,** 406, (6797), 775-781.

6. Fenical, W.; Jensen, P. R., Developing a new resource for drug discovery: marine actinomycete bacteria. *Nature Chemical Biology* **2006,** 2, (12), 666-673.

7. Bender, C. L.; Alarcon-Chaidez, F.; Gross, D. C., Pseudomonas syringae phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and Molecular Biology Reviews* **1999,** 63, (2), 266-+.

8. Grgurina, I.; Barca, A.; Cervigni, S.; Gallo, M.; Scaloni, A.; Pucci, P., Relevance of Chlorine-Substituent for the Antifungal Activity of Syringomycin and Syringotoxin, Metabolites of the Phytopathogenic Bacterium Pseudomonas-Syringae Pv Syringae. *Experientia* **1994,** 50, (2), 130-133.

9. Neumann, C. S.; Fujimori, D. G.; Walsh, C. T., Halogenation strategies in natural product biosynthesis. *Chemistry & Biology* **2008,** 15, (2), 99-109.

10. Ueki, M.; Galonic, D. P.; Vaillancourt, F. H.; Garneau-Tsodikova, S.; Yeh, E.; Vosburg, D. A.; Schroeder, F. C.; Osada, H.; Walsh, C. T., Enzymatic generation of the antimetabolite gamma,gamma-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chemistry & Biology* **2006,** 13, (11), 1183-1191.

11. Fraaije, M. W.; Mattevi, A., Flavoenzymes: diverse catalysts with recurrent features. *Trends in Biochemical Sciences* **2000,** 25, (3), 126-132.

12. Massey, V., Flavoprotein Structure and Mechanism - Introduction. *Faseb Journal* **1995,** 9, (7), 473-475.

13. Thorpe, C.; Kim, J. J. P., Flavoprotein Structure and Mechanism .3. Structure and Mechanism of Action of the Acyl-Coa Dehydrogenases. *Faseb Journal* **1995,** 9, (9), 718-725.

14. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J., Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **2006,** 100, (4), 644-69.

15. Janc, J. W.; Egan, L. A.; Townsend, C. A., Purification and Characterization of Clavaminate Synthase from Streptomyces-Antibioticus - a Multifunctional Enzyme of Clavam Biosynthesis. *Journal of Biological Chemistry* **1995,** 270, (10), 5399-5404.

16. Lloyd, M. D.; Lee, H. J.; Harlos, K.; Zhang, Z. H.; Baldwin, J. E.; Schofield, C. J.; Charnock, J. M.; Garner, C. D.; Hara, T.; van Scheltinga, A. C. T.; Valegard, K.; Viklund, J. A. C.; Hajdu, J.; Andersson, I.; Danielsson, A.; Bhikhabhai, R., Studies on the active site of deacetoxycephalosporin C synthase. *Journal of Molecular Biology* **1999,** 287, (5), 943-960.

17. Ryle, M. J.; Liu, A.; Muthukumaran, R. B.; Ho, R. Y. N.; Koehntop, K. D.; McCracken, J.; Que, L.; Hausinger, R. P., O-2- and alpha-ketoglutarate-dependent tyrosyl radical formation in TauD, an alpha-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **2003,** 42, (7), 1854-1862.

18. Dunwell, J. M.; Culham, A.; Carter, C. E.; Sosa-Aguirre, C. R.; Goodenough, P. W., Evolution of functional diversity in the cupin superfamily. *Trends Biochem Sci* **2001,** 26, (12), 740-6.

19. Hegg, E. L.; Que, L., Jr., The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* **1997,** 250, (3), 625-9.

20. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006,** 440, (7082), 368-71.

21. Hausinger, R. P., FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **2004,** 39, (1), 21-68.

22. Price, J. C.; Barr, E. W.; Hoffart, L. M.; Krebs, C.; Bollinger, J. M., Jr., Kinetic dissection of the catalytic mechanism of taurine:alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2005,** 44, (22), 8138-47.

23. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M., Jr.; Krebs, C., The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2003,** 42, (24), 7497-508.

24. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Jr., Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). *J Am Chem Soc* **2003,** 125, (43), 13008-9.

25. Elkins, J. M.; Ryle, M. J.; Clifton, I. J.; Hotopp, J. C. D.; Lloyd, J. S.; Burzlaff, N. I.; Baldwin, J. E.; Hausinger, R. P.; Roach, P. L., X-ray crystal structure of Escherichia coli

taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* **2002,** 41, (16), 5185-5192.

26. Koehntop, K. D.; Emerson, J. P.; Que, L., Jr., The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J Biol Inorg Chem* **2005,** 10, (2), 87-93.

27. Galonic, D. P.; Barr, E. W.; Walsh, C. T.; Bollinger, J. M., Jr.; Krebs, C., Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat Chem Biol* **2007,** 3, (2), 113-6.

28. Fujimori, D. G.; Barr, E. W.; Matthews, M. L.; Koch, G. M.; Yonce, J. R.; Walsh, C. T.; Bollinger, J. M.; Krebs, C.; Riggs-Gelasco, P. J., Spectroscopic evidence for a high-spin Br-Fe(IV)-Oxo intermediate in the alpha-ketoglutarate-dependent halogenase CytC3 from Streptomyces. *Journal of the American Chemical Society* **2007,** 129, (44), 13408-+.

29. Jeggo, P.; Defais, M.; Samson, L.; Schendel, P., Adaptive Response of Escherichia-Coli to Low-Levels of Alkylating Agent - Comparison with Previously Characterized DNA-Repair Pathways. *Molecular & General Genetics* **1977,** 157, (1), 1-9.

30. Samson, L.; Cairns, J., New Pathway for DNA-Repair in Escherichia-Coli. *Nature* **1977,** 267, (5608), 281-283.

31. Sedgwick, B.; Lindahl, T., Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* **2002,** 21, (58), 8886-8894.

32. Margison, G., A new damage limitation exercise: ironing (Fe(II) out minor DNA methylation lesions. *DNA Repair* **2002,** 1, (12), 1057-1061.

33. Kleibl, K., Molecular mechanisms of adaptive response to alkylating agents in Escherichia coli and some remarks on O-6-methylguanine DNA-methyltransferase in other organisms. *Mutation Research-Reviews in Mutation Research* **2002,** 512, (1), 67-84.

34. Drablos, F.; Feyzi, E.; Aas, P. A.; Vaagbo, C. B.; Kavli, B.; Bratlie, M. S.; Pena-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E., Alkylation damage in DNA and RNA - repair mechanisms and medical significance. *DNA Repair* **2004,** 3, (11), 1389-1407.

35. Landini, P.; Hajec, L. I.; Volkert, M. R., Structure and Transcriptional Regulation of the Escherichia-Coli Adaptive Response Gene Aidb. *Journal of Bacteriology* **1994,** 176, (21), 6583-6589.

36. Landini, P.; Volkert, M. R., Transcriptional Activation of the Escherichia-Coli Adaptive Response Gene Aidb Is Mediated by Binding of Methylated Ada Protein - Evidence for a New Consensus Sequence for Ada-Binding Sites. *Journal of Biological Chemistry* **1995,** 270, (14), 8285-8289.

37. Rohankhedkar, M. S.; Mulrooney, S. B.; Wedemeyer, W. J.; Hausinger, R. P., The AidB component of the Escherichia coli adaptive response to alkylating agents is a flavin-containing, DNA-binding protein. *J Bacteriol* **2006**, 188, (1), 223-30.

38. Aravind, L.; Koonin, E. V., The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* **2001**, 2, (3), RESEARCH0007.

39. Volkert, M. R.; Nguyen, D. C.; Beard, K. C., Escherichia-Coli Gene Induction by Alkylation Treatment. *Genetics* **1986**, 112, (1), 11-26.

40. Frick, L. E.; Massachusetts Institute of Technology. Biological Engineering Division. The versatile E. coli adaptive response protein AlkB mitigates toxicity and mutagenicity of etheno-, ethano-, and methyl-modified bases in vivo. Thesis Ph. D. --Massachusetts Institute of Technology Biological Engineering Division 2007., 2007.

41. Volkert, M. R.; Hajec, L. I.; Nguyen, D. C., Induction of the Alkylation-Inducible Aidb Gene of Escherichia-Coli by Anaerobiosis. *Journal of Bacteriology* **1989**, 171, (2), 1196-1198.

42. Matijasevic, Z.; Hajec, L. I.; Volkert, M. R., Anaerobic Induction of the Alkylation-Inducible Escherichia-Coli Aidb Gene Involves Genes of the Cysteine Biosynthetic-Pathway. *Journal of Bacteriology* **1992**, 174, (6), 2043-2046.

43. Smirnova, G. V.; Oktyabrsky, O. N.; Moshonkina, E. V.; Zakirova, N. V., Induction of the Alkylation-Inducible Aidb Gene of Escherichia-Coli by Cytoplasmic Acidification and N-Ethylmaleimide. *Mutation Research* **1994**, 314, (1), 51-56.

44. Tiffany, K. A.; Roberts, D. L.; Wang, M.; Paschke, R.; Mohsen, A. W. A.; Vockley, J.; Kim, J. J. P., Structure of human isovaleryl-CoA dehydrogenase at 2.6 angstrom resolution: Basis for substrate specificity. *Biochemistry* **1997**, 36, (28), 8455-8464.

45. Bowles, T.; Metz, A. H.; O'Quin, J.; Wawrzak, Z.; Eichman, B. F., Structure and DNA binding of alkylation response protein AidB. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, (40), 15299-15304.

46. Wei, Y. F.; Carter, K. C.; Wang, R. P.; Shell, B. K., Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Research* **1996**, 24, (5), 931-937.

47. Westbye, M. P.; Feyzi, E.; Aas, P. A.; Vagbo, C. B.; Talstad, V. A.; Kavli, B.; Hagen, L.; Sundheim, O.; Akbari, M.; Liabakk, N. B.; Slupphaug, G.; Otterlei, M.; Krokan, H. E., Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *Journal of Biological Chemistry* **2008**, 283, (36), 25046-25056.

48. Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine and 3-methylthymine, in alkB Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, (39), 14051-14056.

49. Koivisto, P.; Robins, P.; Lindahl, T.; Sedgwick, B., Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *Journal of Biological Chemistry* **2004,** 279, (39), 40470-40474.

50. Delaney, J. C.; Smeester, L.; Wong, C. Y.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M., AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nature Structural & Molecular Biology* **2005,** 12, (10), 855-860.

51. Frick, L. E.; Delaney, J. C.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Alleviation of 1,N-6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, (3), 755-760.

52. Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B., Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** 99, (26), 16660-16665.

53. Aas, P. A.; Otterlei, M.; Falnes, P. O.; Vagbo, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjoras, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **2003,** 421, (6925), 859-863.

54. Ringvoll, J.; Moen, M. N.; Nordstrand, L. M.; Meira, L. B.; Pang, B.; Bekkelund, A.; Dedon, P. C.; Bjelland, S.; Samson, L. D.; Falnes, P. O.; Klungland, A., AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Research* **2008,** 68, (11), 4142-4149.

55. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F., Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **2006,** 439, (7078), 879-884.

56. Yang, C. G.; Yi, C. Q.; Duguid, E. M.; Sullivan, C. T.; Jian, X.; Rice, P. A.; He, C., Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **2008,** 452, (7190), 961-U4.

57. Sundheim, O.; Vagbo, C. B.; Bjoras, M.; Sousa, M. M. L.; Talstad, V.; Aas, P. A.; Drablos, F.; Krokan, H. E.; Tainer, J. A.; Slupphaug, G., Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *Embo Journal* **2006,** 25, (14), 3389-3397.

58. Keller, S.; Wage, T.; Hohaus, K.; Holzer, M.; Eichhorn, E.; van Pee, K. H., Purification and partial characterization of tryptophan 7-halogenase (PrnA) from Pseudomonas fluorescens. *Angewandte Chemie-International Edition* **2000,** 39, (13), 2300-2302.

59. Yeh, E.; Garneau, S.; Walsh, C. T., Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin

biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, (11), 3960-3965.

60. Zehner, S.; Kotzsch, A.; Bister, B.; Sussmuth, R. D.; Mendez, C.; Salas, J. A.; van Pee, K. H., A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in Streptomyces rugosporus LL-42D005. *Chemistry & Biology* **2005**, 12, (4), 445-452.

61. Dorrestein, P. C.; Yeh, E.; Garneau-Tsodikova, S.; Kelleher, N. L.; Walsh, C. T., Dichlorination of a pyrrolyl-S-carrier protein by FADH(2)-dependent halogenase PltA during pyoluteorin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, (39), 13843-13848.

62. Buedenbender, S.; Rachid, S.; Muller, R.; Schulz, G. E., Structure and Action of the Myxobacterial Chondrochloren Halogenase CndH: A New Variant of FAD-dependent Halogenases. *Journal of Molecular Biology* **2009**, 385, (2), 520-530.

63. Yeh, E.; Cole, L. J.; Barr, E. W.; Bollinger, J. M.; Ballou, D. P.; Walsh, C. T., Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry* **2006**, 45, (25), 7904-7912.

64. Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T., Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **2007**, 46, (5), 1284-92.

65. Vaillancourt, F. H.; Yin, J.; Walsh, C. T., SyrB2 in syringomycin E biosynthesis is a nonheme Fe-II alpha-ketoglutarate- and O-2-dependent halogenase. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, (29), 10111-10116.

66. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006**, 440, (7082), 368-371.

67. Bush, J. A.; Long, B. H.; Catino, J. J.; Bradner, W. T., Production and Biological-Activity of Rebeccamycin, a Novel Antitumor Agent. *Journal of Antibiotics* **1987**, 40, (5), 668-678.

68. Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T., Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **2005**, 436, (7054), 1191-1194.

69. Kakeya, H.; Zhang, H. P.; Kobinata, K.; Onose, R.; Onozawa, C.; Kudo, T.; Osada, H., Cytotrienin A, a novel apoptosis inducer in human leukemia HL-60 cells. *Journal of Antibiotics* **1997**, 50, (4), 370-372.

70. Neidig, M. L.; Brown, C. D.; Light, K. M.; Fujimori, D. G.; Nolan, E. M.; Price, J. C.; Barr, E. W.; Bollinger, J. M.; Krebs, C.; Walsh, C. T.; Solomon, E. I., CD and MCD of CytC3 and taurine dioxygenase: Role of the facial triad in alpha-KG-dependent oxygenases. *Journal of the American Chemical Society* **2007**, 129, (46), 14224-14231.

71. Fujimori, D. G.; Walsh, C. T., What's new in enzymatic hallogenations. *Current Opinion in Chemical Biology* **2007,** 11, (5), 553-560.

72. Lloyd, M. D.; Lipscomb, S. J.; Hewitson, K. S.; Hensgens, C. M. H.; Baldwin, J. E.; Schofield, C. J., Controlling the substrate selectivity of deacetoxycephalosporin/deacetylcephalosporin C synthase. *Journal of Biological Chemistry* **2004,** 279, (15), 15420-15426.

I.K. Tables, Reaction Schemes and Figures

Table I.1. List of substrate preferences for AlkB and two human homologues, hABH2 and hABH3.

	DNA or RNA preference	Lesions preferred
AlkB	dsDNA, ssDNA, RNA	1mA, 3mC, 3mT, 1mG, εA, EA
hABH2	dsDNA	1mA
hABH3	ssDNA, RNA	1mA, 3mC

Table I.2. Steady-state kinetics of CytC1 for different amino acid substrates.[10] Amino acid adenylation by CytC1 is a reversible reaction. The kinetics are measured through the amino acid-dependent exchange of [32]PPi into ATP.

Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} mM^{-1})
L-Val	0.60 ± 0.05	17.2 ± 0.4	29 ± 3
1-aminocyclopropane-1-carboxylate	1.65 ± 0.07	11.9 ± 0.2	7.2 ± 0.4
D-Aba	4.0 ± 0.3	17.4 ± 0.5	4.3 ± 0.4
D-Val	6.9 ± 0.5	10.7 ± 0.4	1.6 ± 0.2
4-Cl-L-Aba	7.6 ± 0.6	7.6 ± 0.3	1.0 ± 0.1
L-Aba	13 ± 1	10.7 ± 0.5	0.8 ± 0.1
L-*allo*-Ile	18 ± 3	12 ± 1	0.6 ± 0.2
D-Cys	47 ± 12	29 ± 6	0.6 ± 0.3
L-Cys	5.2 ± 0.7	2.7 ± 0.2	0.5 ± 0.1

Scheme I.1. Mechanism of dehydrogenation reaction as seen in acyl CoA dehydrogenase.[13]
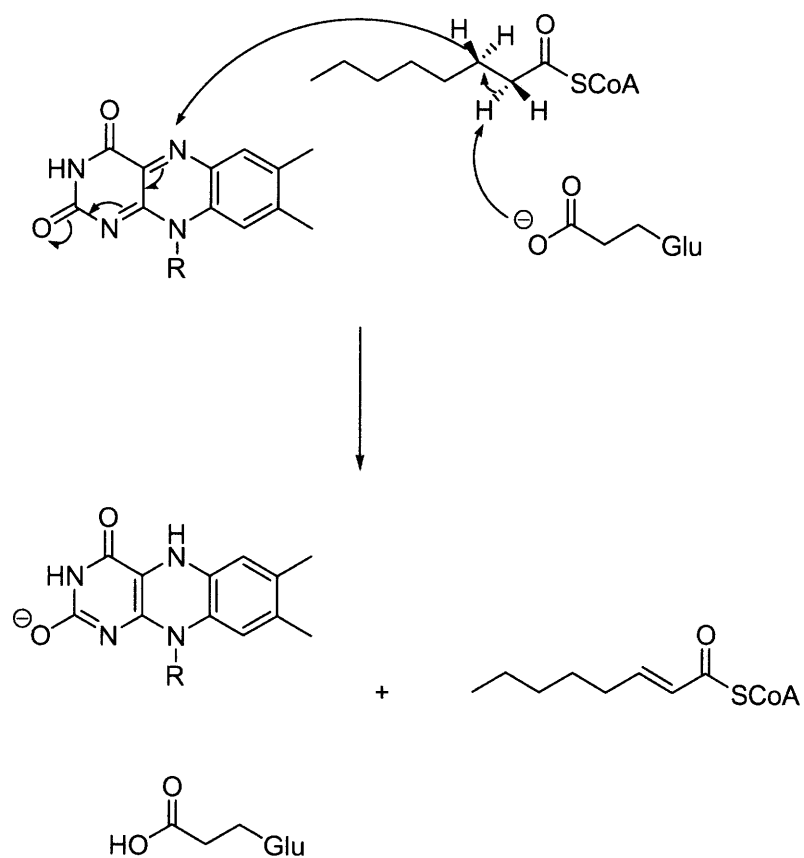
Scheme I.2. FAD chemistry in flavin-dependent hydroxylation versus halogenation.[12, 71]

FADH$_2$

+ O$_2$

C4a-hydroperoxy FAD

Hydroxylation

Substrate

Substrate-OH

Halogenation

Cl$^-$

HOCl

C4a-hydroxy FAD

H$_2$O

FAD
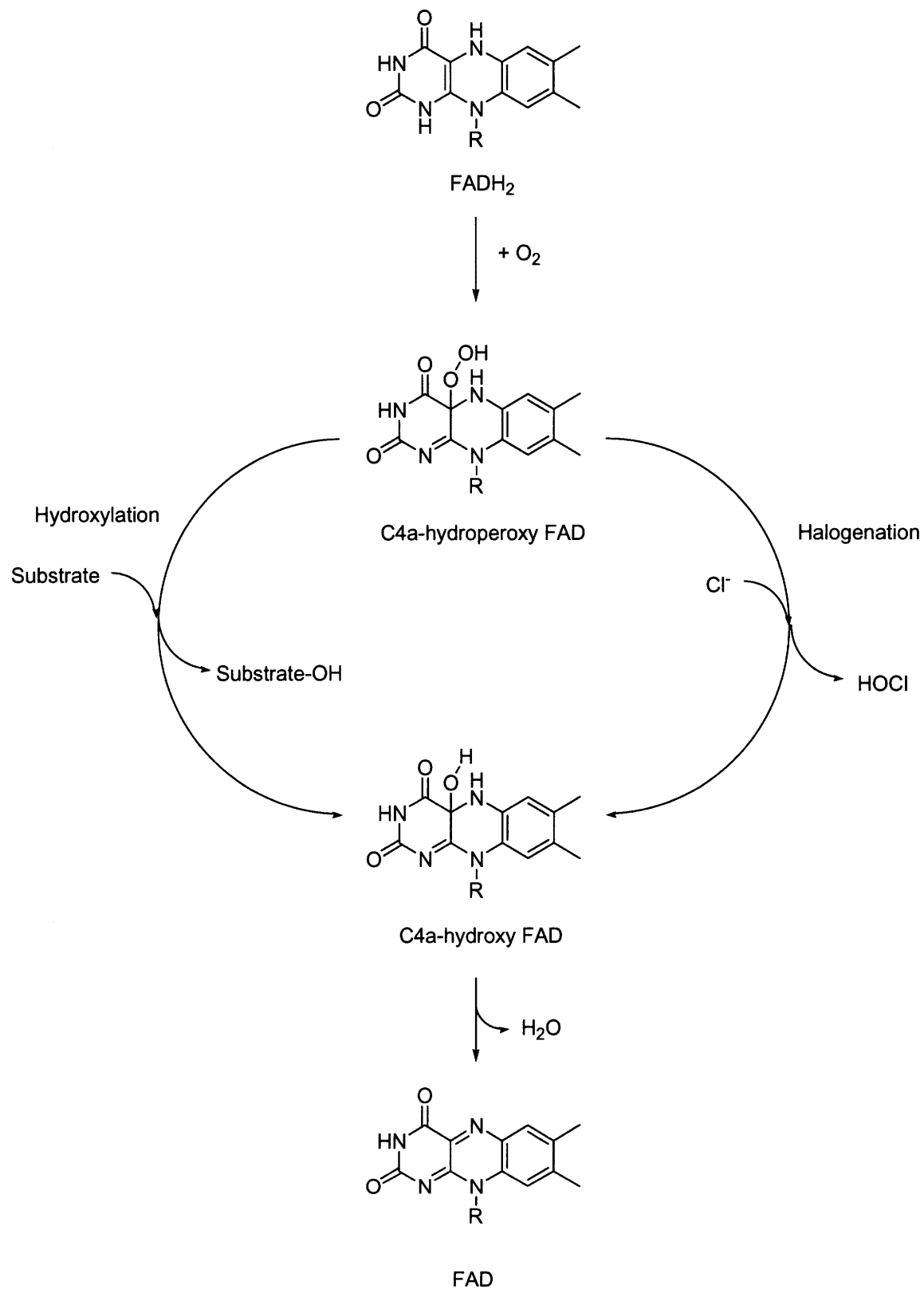
Scheme I.3. Reaction scheme of enzymes in the αKG/Fe(II)-dependent superfamily. a) clavaminate synthase (CAS). CAS catalyzes the hydroxylation of deoxyguanidinoproclavaminic acid, as well as the cyclization and desaturation reactions of proclavaminic acid in the synthesis of clavulanic acid.[15] b) Reaction scheme of deacetoxycephalosporin C synthase (DAOCS). DAOCS catalyzes the ring expansion and hydroxylation of penicillin N in the synthesis of cephalosporin.[72] c) Reaction scheme of taurine dioxygenase (TauD). TauD catalyzes the hydroxylation of taurine.[17] d) Reaction scheme of chlorination of L-aminobuytric acid attached to a stand-alone thiolation domain CytC2 by the non-heme iron halogenase CytC3. The thiolation domain CytC2 is labeled as T, and the phosphopantetheine arm that is post-translationally attached to the thiolation domain is indicated by a wavy line.

Scheme II.3.
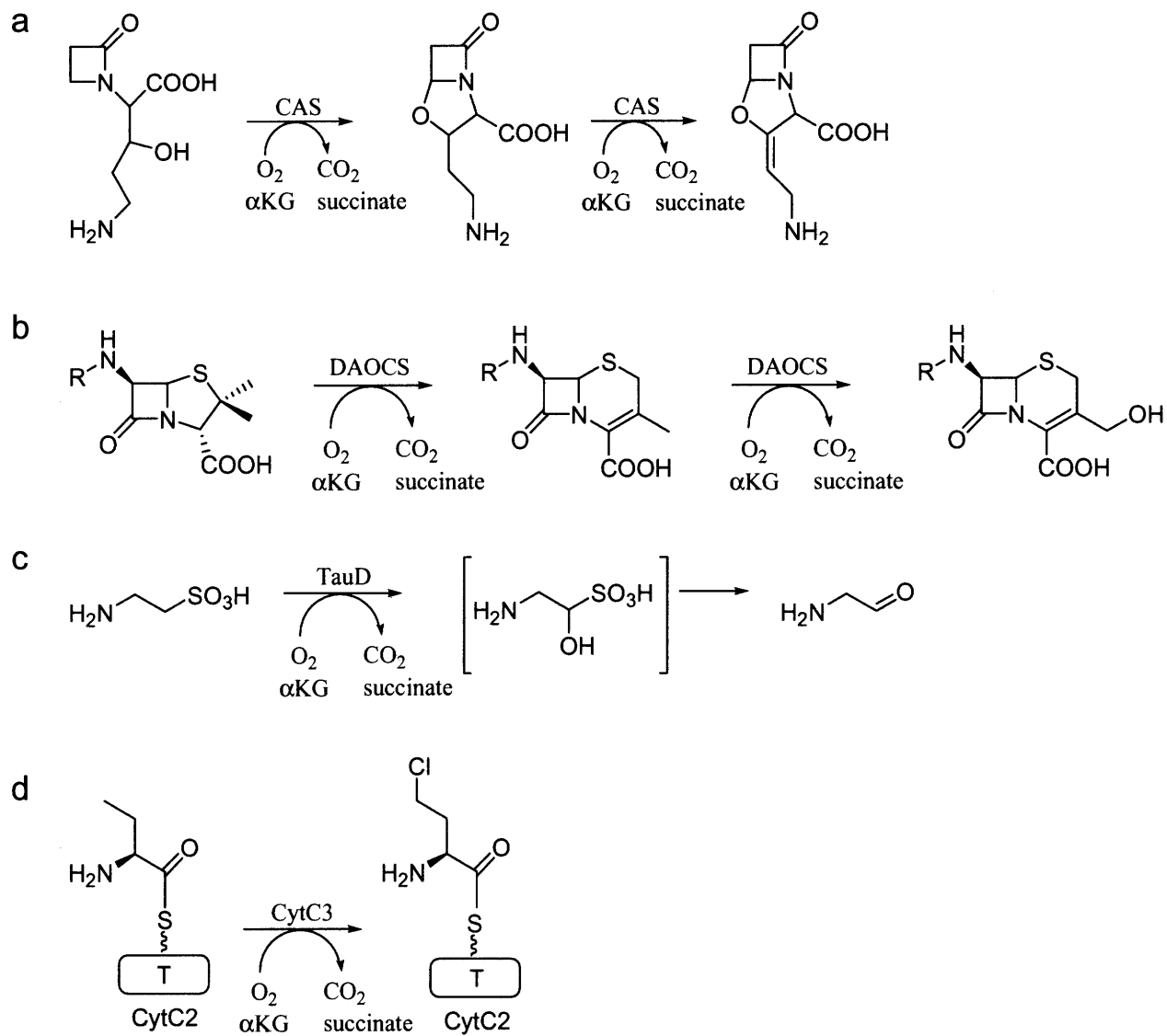
Scheme I.4. Reaction scheme of AlkB with 1mA and 3mC substrate.
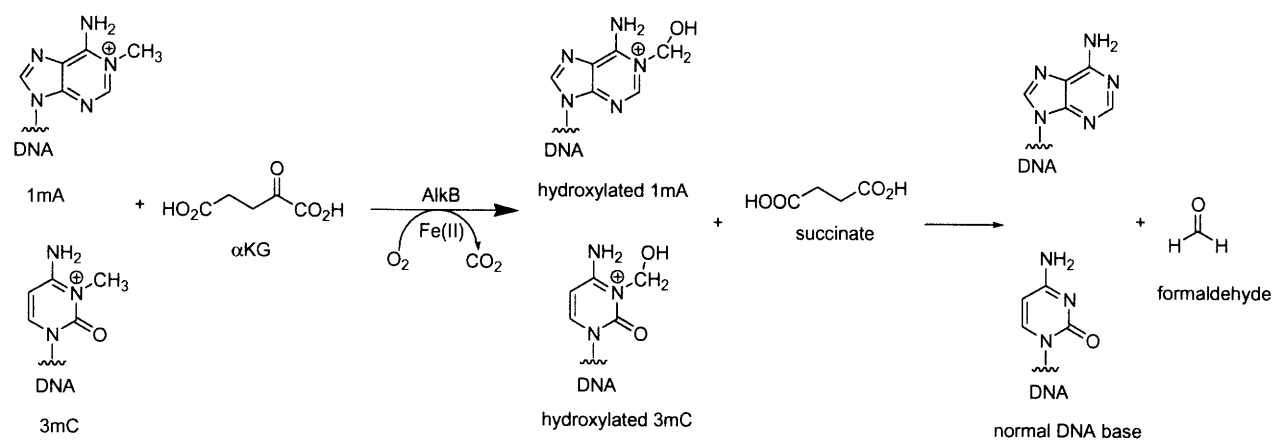
Scheme I.5. Reaction scheme of AidB proposed by Hausinger *et al.*[37] AidB may repair DNA

lesions through dehydrogenase activity, forming a double bond between the methyl group and

the normal base. The Schiff's base can be hydrolyzed by water to leave as formaldehyde.

$$
\begin{array}{c}
\text{DNA} \\
\mid \\
H_3C-NH(\text{base})
\end{array}
\longrightarrow
\begin{array}{c}
\text{DNA} \\
\mid \\
H_2C=N(\text{base})
\end{array}
\longrightarrow
\begin{array}{c}
\text{DNA} \\
\mid \\
NH_2(\text{base})
\end{array}
$$

AidB-FAD   AidB-FADH$_2$ H$_2$O   HCHO

H$_2$O$_2$ O$_2$
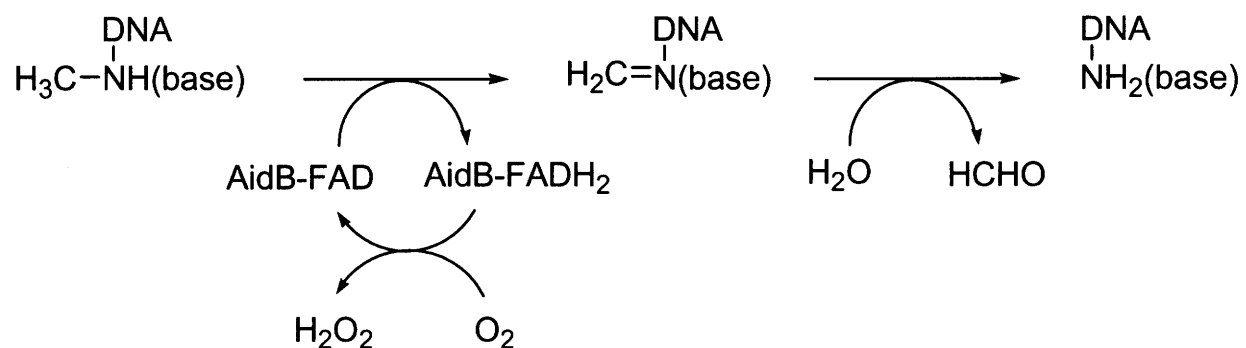
Scheme I.6. Flavin-dependention halogenations reactions. a) Regioselective chlorination of tryptophan by FAD-dependent halogenases, PrnA or RebH (top of the reaction scheme), PyrH (middle), and Thal (bottom).[9] b) Reaction scheme of dichlorination of the pyrrole moiety of the oxidized proline attached to a thiolation domain by the flavin-dependent halogenase PltA. The thiolation domain is labeled as T, and the phosphopantetheine arm that is post-translationally attached to the thiolation domain is indicated by a wavy line.
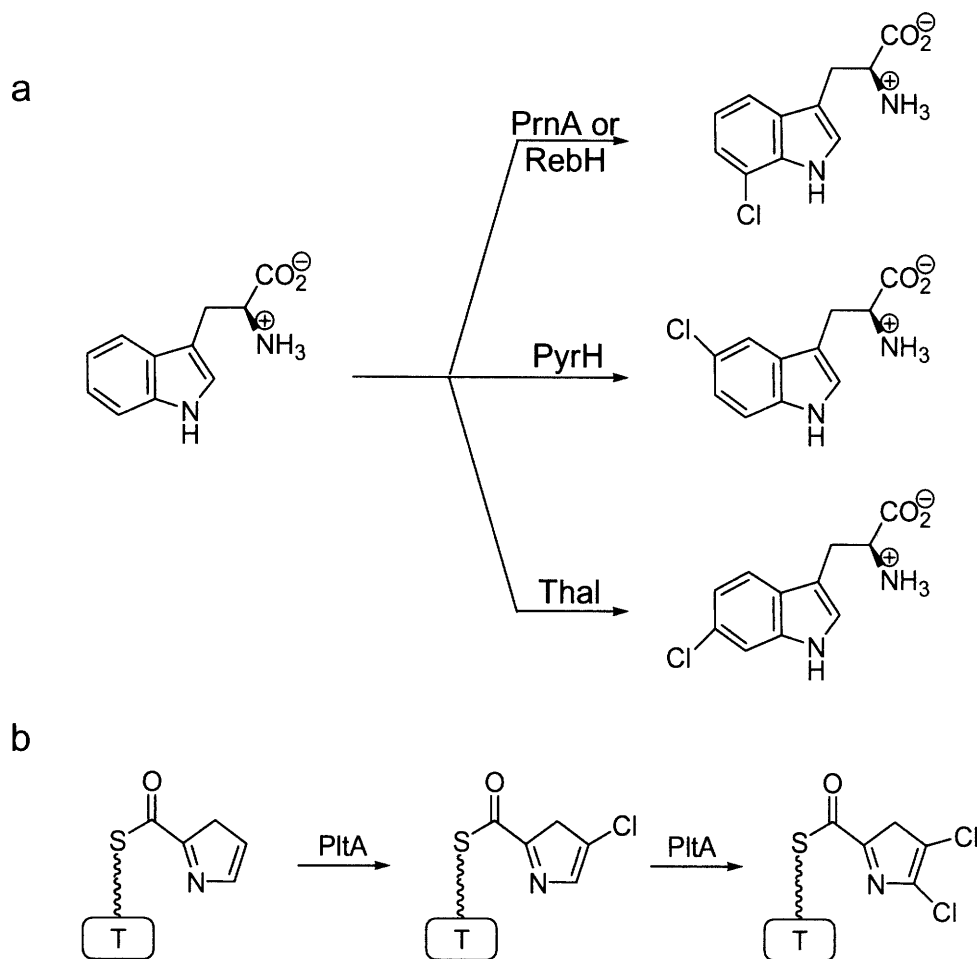
Scheme I.7. Reaction scheme for the chlorination of L-threonine attached to the thiolation domain of SyrB1 by the non-heme iron halogenase SyrB2.[65] SyrB1 consists of an adenylation domain and a thiolation domain, labeled as A and T, respectively. The phosphopantetheine arm that is post translationally linked to the thiolation domain is indicated by a wavy line.
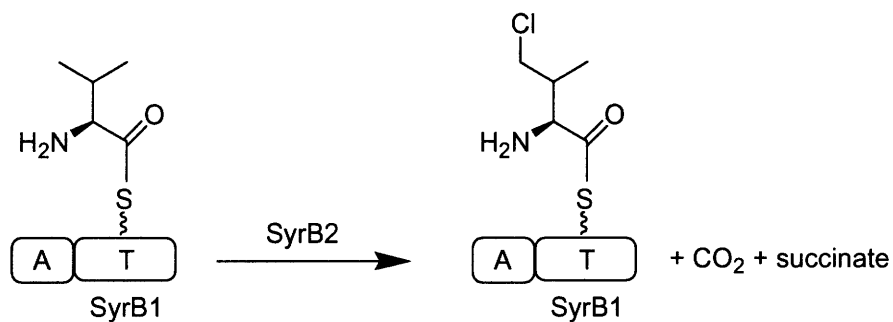
Scheme I.8. Production of 4-Cl-L-Aba by CytC1-3. CytC1 loads L-Aba-AMP onto the phosphopantetheine arm (wavy line) on CytC2. CytC3 chlorinates the tethered substrate to form 4-Cl-L-Aba.[10] The 4-Cl-L-Aba is postulated to form Cytotreinin A.[68] CytC1 is an adenylation domain (A), and CytC2 is a thiolation domain (T).

Cytotreinin A
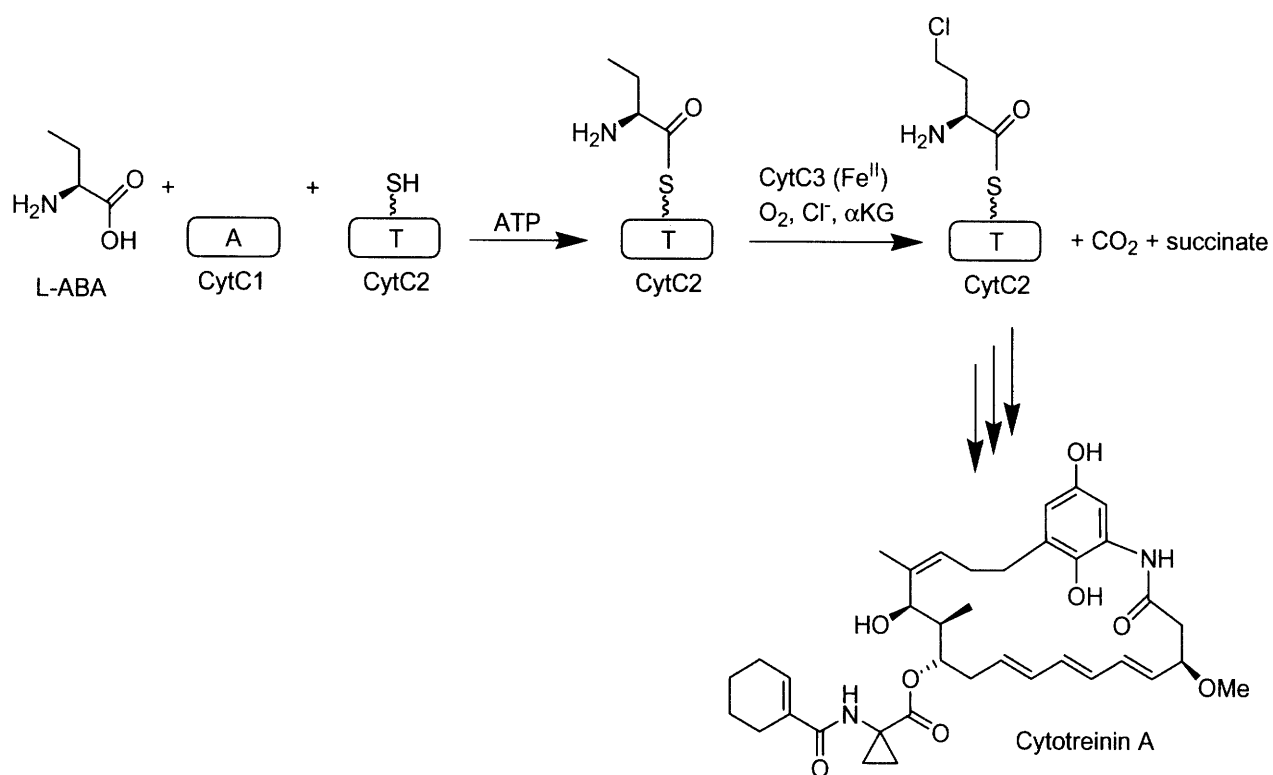
Figure I.1. The conserved "jelly-roll" or β-barrel motif in all αKG/Fe(II)-dependent superfamily members. The conserved 2-His-1-carboxylate motif that coordinates the iron is shown as sticks, with iron as a sphere.

Figure I.2. General mechanism of hydroxylation by αKG/Fe(II)-dependent enzymes.[21]
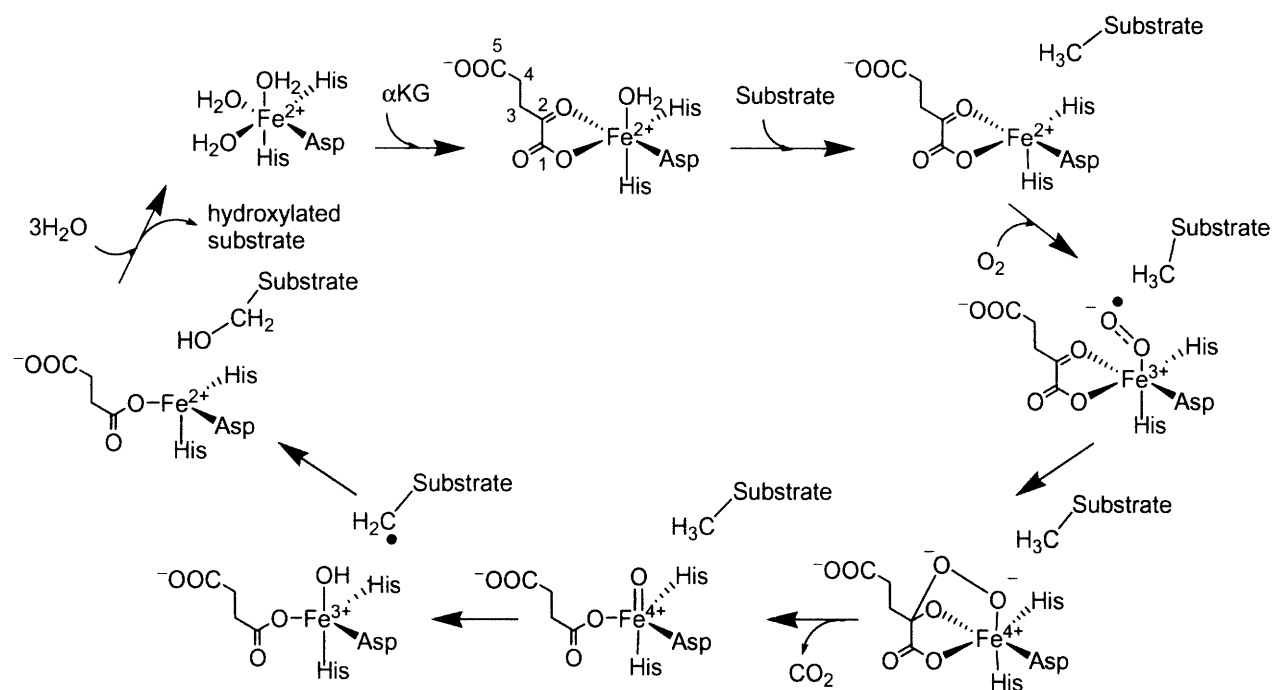
Figure I.3. Sites of DNA methylation. 12 sites are on DNA bases, and one is on the phosphate backbone of DNA. Red arrows indicate DNA methylation sites in single stranded DNA (ssDNA) that are caused by S_N2 methylating agents. Blue arrows indicate DNA methylation sites that are caused by S_N1 methylating agents. Green arrows indicate DNA sites that are alkylated by most alkylating agents. The yellow arrow indicates the methylation sites created by reaction with a methyl radical. The percentage near each arrow indicates the relative abundance modifications at that site in DNA.[1]
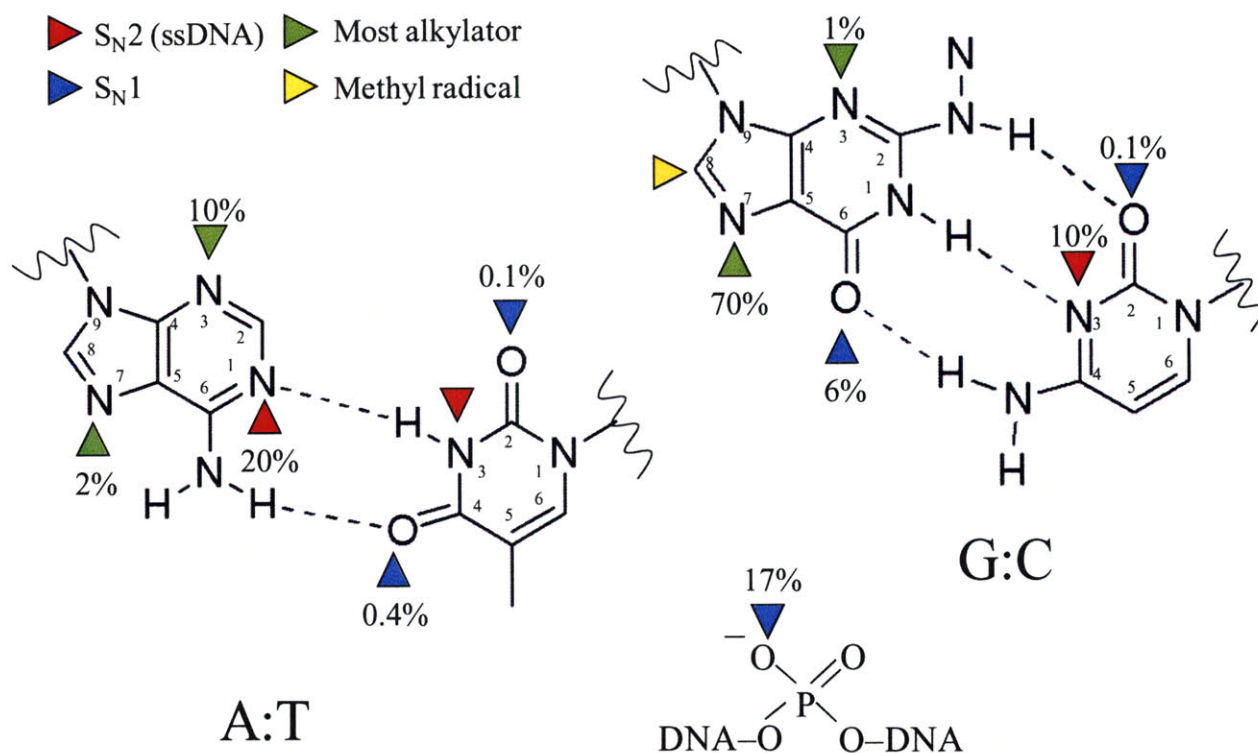
Figure I.4. Schematic diagram of the adaptive response in *E. coli*. The Ada protein acts as a transcription factor when its N-terminal domain demethylates the phosphate backbone lesion in DNA. The Ada transcription factor binds to the promoter and up-regulates the expression of Ada, AlkB, AlkA, and AidB proteins.
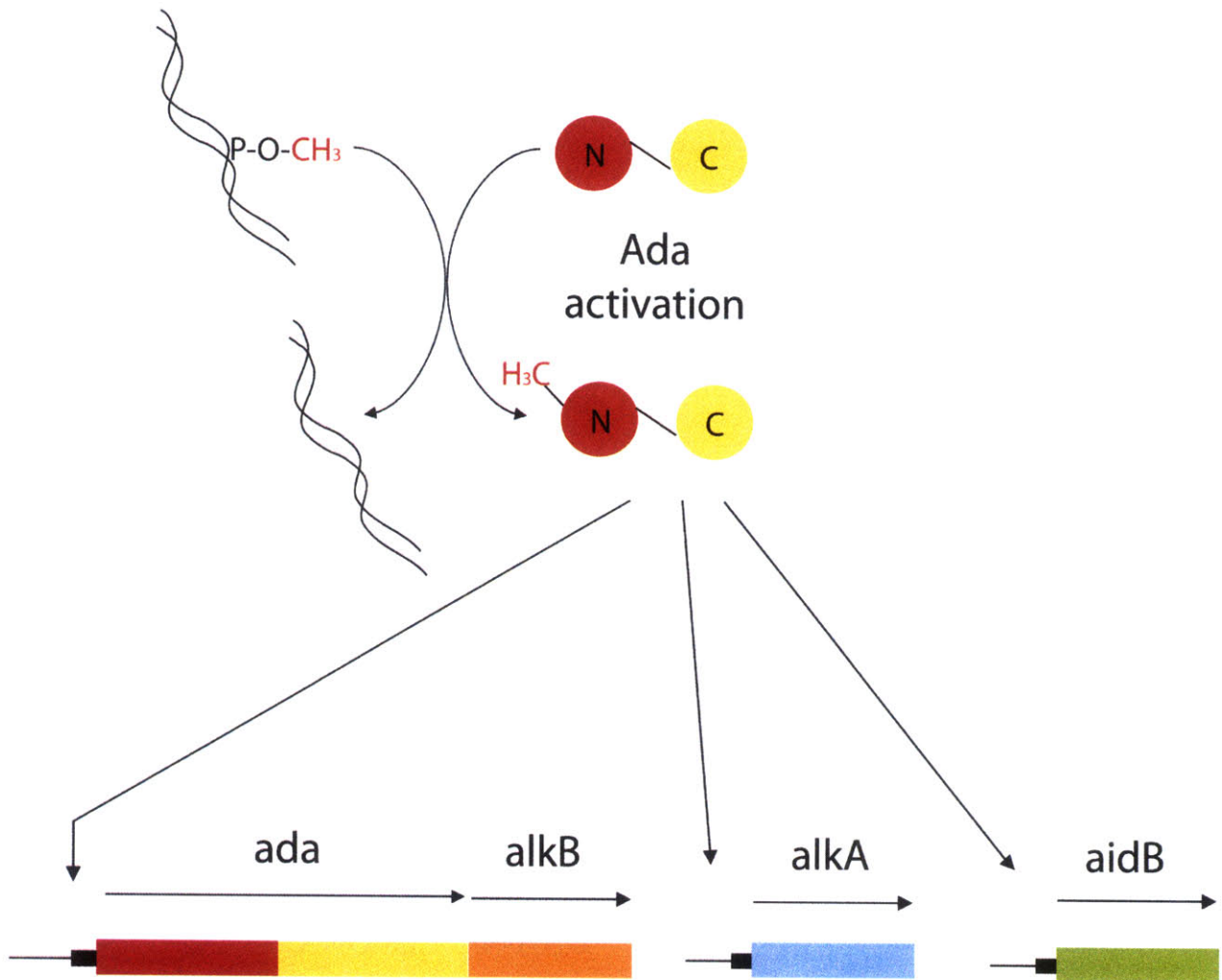
Figure I.5. Sequence alignment of AidB with isovaleryl CoA dehydrogenase (IVD) from human and medium chain acyl CoA dehydrogenase (MCAD) from human. Strictly conserved residues are highlighted in red. Numbers given above the sequence follow the AidB numbering.

```
              1        10        20        30        40        50        60
              .         .         .         .         .         .         .
AidB   .......VHWQTHTVFNQPIPLNNSNLYLSDGALCEAVTREGAGWDSDFLASIGQQLGTAESLELGRLAN
MCAD   .................................GFSFELTEQQKEFQATARKFARE.EIIPVAAEYD
IVD    MATATRLLGWRVASWRLRPPLAGFVSQRAHSLLPVDDAINGLSEEQRQLRQTMAKFLQE.HLAPKAQEID


              70        80        90       100       110       120
              .         .         .         .         .         .
AidB   VNPPELLRYDAQGRRLDDVRFHPAWHLLMQALCTNRVHNLAWEE.....DARSGAFVARAARFMLHAQVE
MCAD   RTGE................YPVPLLKRAWELGLMNTHIPESFGGLGL...GIIDSCLITEELAYGC
IVD    RSNE................FKNLREFWKQLGNLGVLGITAPVQYGGSGL...GYLEHVLVMEEISRAS


              130       140       150       160       170       180       190
              .         .         .         .         .         .         .
AidB   AGSLCPITMTF.AATPLLL.QMLPAPFQDWTTPLLSDRYDSHLLPGGQKRGLLIGMGMTEKQGGSDVMSN
MCAD   TGVQTAIEANT.LGQVPLIIGGNYQQQKKYLGRMTE.............EPLMCAYCVTEPGAGSDVAGI
IVD    GAVGLSYGAHSNLCINQLVRNGNEAQKEKYLPKLIS............GEYIGALAMSEPNAGSDVVSM


              200       210       220       230       240       250
              .         .         .         .         .         .
AidB   TTRAERLEDGSYRLVGHKWFFSVP.QSDAHLVLAQTAG.......GLSCFFVPRFLPDGQRNAIRLERL
MCAD   KTKAEKKGD.EYIINGQKMWITNGGKANWYFLLARSDPDPKAPASKAFTGFIVEAD.....TPGVQIGRK
IVD    KLKAEKKGN.HYILNGNKFWITNGPDADVLIVYAKTDLAAVPAS.RGITAFIVEKG.....MPGFSTSKK


              260       270       280       290       300       310       320
              .         .         .         .         .         .         .
AidB   KDKLGNRSNASCEVEFQDAIG...WLLGLEGEGIRLILKMGGMTRFDCALGSHAMMRRAFSLAIYHAHQR
MCAD   EINMGQRCSDTRGIVFEDVRVPKENVLTGEGAGFKIAMGTFDKTRPPVAACAVGLAQRALDEATKYALER
IVD    LDKLGMRGSNTCELIFEDCKIPAANILGHENKGVYVLMSGLDLERLVLAGGPLGLMQAVLDHTIPYLHVR


              330       340       350       360       370       380       390
              .         .         .         .         .         .         .
AidB   HVFGNPLIQQPLMRHVLSRMALQLEGQTALLFRLARAWDRRADAKEALWARLFTPAAKFVICKRGMPFVA
MCAD   KTFGKLLAEHQGISFLLADMAMKVELARLSYQRAAWEIDSGR......RNTYYASIAKAYAADIANQLAT
IVD    EAFGQKIGHFQLMQGKMADMYTRLMACRQYVYNVAKACDEGH......CTAKDCAGVILYSAECATQVAL


              400       410       420       430       440       450       460
              .         .         .         .         .         .         .
AidB   EAMEVLGGIGYCEESELPRLYREMPVNSIWEGSGNIMCLDVLRVLNKQAGVYDLLSEAFVEVKGQDRYFD
MCAD   DAVQVFGGNGFNTEYPVEKLMRDAKIYQIYEGTAQIQRIIIAREHIGRYK...................
IVD    DGIQCFGGNGYINDFPMGRFLRDAKLYEIGAGTSEVRRLVIGRAFNADFH..................


              470       480       490       500       510       520       530
              .         .         .         .         .         .         .
AidB   RAVRRLQQQLRKPAEELGREITHQLFLLGCGAQMLKYASPPMAQAWCQVMLDTRGGVRLSEQIQNDLLLR
MCAD   ...................................................................
IVD    ...................................................................


              540
              .
AidB   ATGGVCV
MCAD   .......
IVD    .......
```

53

Figure I.6. A model proposed by Eichman *et al.* of AidB protecting DNA from alkylating agents by secluding the DNA helix in a 25 Å pore formed by arranging four tetrameric units around the DNA.[45] One tetrameric unit of AidB is colored by subunit (cyan, green, magenta, and orange), in cartoon representation. The other AidB tetramer units are shown in black cartoon. FADs in the central subunit are shown as yellow spheres. The proposed DNA binding regions of each tetramer AidB unit are colored in pink. DNA helices are shown as light grey circles.

Chapter II

Crystallographic and Biophysical Characterization

of Adaptive Response Flavoprotein AidB

II.A. Summary

AidB is an FAD-binding protein that is induced in the presence of alkylating agents in *E. coli*.[1] AidB shares sequence homology with the acyl CoA dehydrogenase family.[1] Although its exact function is not yet known, it has been shown that AidB has low isovaleryl CoA dehydrogenase (IVD) activity and binds DNA.[1,2] Based on the low IVD activity and its ability to bind DNA, it has been proposed that AidB could repair alkylated DNA lesions by dehydrogenase activity.[1] A recent crystal structure of AidB was determined by Eichman *et al.*, and based on this structure, authors proposed that AidB protects DNA from damaging agents by sequestering the DNA in a large complex formed by the protein with pores that can bind DNA.[3] Both the AidB sequence homology to IVD and previous crystallographic analysis suggest that AidB is a tetramer, just like other enzymes in the acyl CoA dehydrogenase family, but with a different tetrameric arrangement. Here, the crystal structure of AidB was solved in a different space group, $P3_2$, with a different crystal packing. The large complex that was proposed to sequester DNA previously reported was not observed in this crystal form. Furthermore, biophysical data and these crystallographic analyses suggest that the oligomeric state of AidB might be concentration dependent, and thus AidB might function as a dimer depending on the concentration of AidB in the cell.

Protein purification, biophysical studies, crystallization, and structure determination of AidB were done by C. Wong. Structural analysis was done by, and the paper was written by both C. Wong and C.L. Drennan.

II.B. Introduction

DNA damage arises from both endogenous and exogenous alkylating agents, causing cellular cytotoxicity and mutagenicity. In *E. coli*, four enzymes are up-regulated as part of the so-called adaptive response in the presence of alkylating agents to prevent such cellular cytotoxicity and mutagenicity: Ada, AlkA, AlkB and AidB proteins.[4] Ada is the DNA methyltransferase, and upon methylation of its N-terminus, it becomes a transcription factor that regulates the expression level of itself and the adaptive response proteins.[4] AlkA is a DNA glycosylase that repairs lesions that form destabilized glycosylic bonds, including 3-methyladenine, 7-methyladenine, and 7-methylguanine.[5] AlkB is a DNA dioxygenase that mainly repairs lesions that interrupt base pairing, including 1-methyladenine, 3-methylcytocine, 1-N^6-ethenoadenine, and 1-N^6-ethanoadenine.[6-9] AidB is a flavoprotein that is up-regulated in the presence of alkylating agent N-methyl-N'-nitro-N-nitrosoguanidine (MNNG);[10] however, the function of AidB in DNA repair has not yet been identified beyond its ability to nonspecifically bind DNA.[1]

Sequence alignment predicts that AidB belongs to the acyl CoA dehydrogenase superfamily. In addition, AidB possesses low levels of isovaleryl CoA dehydrogenase (IVD) activity.[2] IVD catalyzes the α,β-oxidation of 3-methylbutanoyl CoA to 3-methylcrotonyl CoA as part of the leucine metabolism pathway. IVD is a homotetramer in solution and contains one FAD per monomer. Biochemical studies show AidB also contains FAD. However, whether the role of FAD in AidB is structural or catalytic remains unclear because no true function for AidB has been identified. On the basis of AidB's IVD activity, Hausinger *et al.* have proposed a DNA repair reaction for AidB in which AidB uses dehydrogenase activity to demethylate a damaged lesion. In particular, AidB could use dehydrogenase activity to form a double bond between the

methyl group and the base, which could then be released as formaldehyde via a hydrolysis reaction (Scheme II.1).[1]

A recent structural study reports that AidB is a homotetramer, with FAD bound at the interface of two monomers.[3] Based on the crystal structure packing and a biochemical study showing by gel shift assays that AidB binds DNA,[1] Eichman *et al.* have proposed that double-stranded DNA (dsDNA) binds to AidB at two ends of the homotetramer, and that tetramer units can arrange themselves into a "mega complex" with pores that are 25 Å in diameter for sequestering DNA (Figure II.1).[3] Here we present the crystal structure of AidB that has different crystal packing, along with biophysical data suggesting that the oligomeric state of AidB might be concentration-dependent. We further consider the relevance of oligomeric state to the proposed functions for AidB. To address whether the FAD is likely to have a catalytic or structural role, we use site-directed mutagenesis and analytical ultracentrifugation to determine if FAD binding is needed for AidB dimerization. We also analyze the structure of FAD binding site. This new crystal form of AidB with the biochemical data presented here suggest that AidB does more than merely shielding DNA from chemical attack.

II.C. Materials and Methods

Cloning and Over Expression of AidB

The AidB gene was amplified from host *E. coli* AB1157 strain using the forward primer AidB-Nco-F 5'-AGGATATA<u>CCATGG</u>AGGGAGACACAGTGCAC-3' (the NcoI restriction enzyme recognition site is underlined) and the reverse primer AidB-Hind-R 5'-TCTCGTAG<u>AAGCTT</u>TTACACACACACTCCCCCCG-3' (the HindIII restriction enzyme recognition site is underlined). The PCR amplification used annealing temperature of 50°C and a

5 min extension time over 20 cycles. The amplified gene was digested with NcoI and HindIII (New England Biolab) and subcloned into a pET28a vector (Novagen) that had also been digested with NcoI and HindIII. The AidB-pET28a clone was confirmed to be correct by DNA sequence analysis, and was transformed into *E. coli* BL21(DE3) cells for expression. Overnight cultures were grown in Luria-Bertani (LB) media with kanamycin (50 μg/mL) at 37°C with constant shaking at 250 rpm. These overnight cultures were used to inoculate 8 L of LB media containing kanamycin (50 μg/mL). Cells were grown at 37°C to an optical density at 600 nm of ~0.5, at which point isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM for induction, and the cultures were transferred to 21°C for overnight growth. Cells were harvested by centrifugation at 6000 x g for 20 min, cell pellets were frozen in liquid nitrogen and stored at -80°C until purification.

Various AidB mutants were made by site-directed mutagenesis of the wild type AidB-pET28a clone. AidB mutants were over expressed in the same condition as wild type described above. The primers used for each AidB mutants are listed in Table II.1.

Purification of AidB

For a typical 8 L cell culture, 25 g of wet cells were resuspeded in 125 mL lysis buffer (10 mM Tris, pH 7.8, 1 mM ethylenediaminetetraacetic acid (EDTA), 300 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol (BME)), homogenized by sonication, and the cell debris was separated from the soluble supernatant by centrifugation at 35,000 x g for 30 min at 4°C. The cell lysate was treated with 30% and 45% ammonium sulfate precipitation at room temperature, with each addition followed by centrifugation at 6,000 x g at 4°C to collect the precipitated protein. The protein that precipitated between 30 to 45% ammonium sulfate was saved and

dissolved in lysis buffer with 1 M NaCl. The soluble protein was loaded onto an S200 size exclusion fast performance liquid chromatography column (EMD Biosciences) that was equilibrated with lysis buffer, and the protein was eluted with 1.5 column volumes of lysis buffer. Fractions containing AidB were pooled and loaded onto a phenyl-agarose column (EMD Bioscience) equilibrated with lysis buffer. The column was washed with 10 column volumes of lysis buffer, and the protein was eluted with lysis buffer containing 0.4% w/v deoxycholate. The protein was dialyzed against storage buffer (10 mM Tris, pH 7.8, 100 mM NaCl, 10% glycerol and 2 mM BME) overnight at 4°C. The purified AidB was analyzed by SDS-PAGE (Figure II.2). AidB protein was concentrated to ~20 mg/mL and flash frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined by measuring the absorbance at 280 nm, and the theoretical extinction coefficient (76,150 $M^{-1}cm^{-1}$) was used for calculating the concentration. A typical 8 L culture yielded about 2 – 3 mL of 20 mg/mL purified protein. Wild type AidB and various mutants were purified the same way.

Crystallization, Data Collection and Processing

AidB crystallized in several different conditions using the hanging drop method by mixing 1 μL of protein (10 mg/mL AidB in 10 mM Tris, pH 7.8, 100 mM NaCl, 10% glycerol and 2 mM BME) with 1 μL of precipitant solution (Hampton Research): 0.1 M HEPES, pH 7.5, 20% v/v ethanol, 0.25 M MgCl$_2$ (Figure II.3.a); 0.1 M Tris, pH 8.0, 2% v/v tacsimate, 20% w/v PEG 3,000 (Figure II.3.b); and 0.1 M imidazole, pH 8.0, 0.3 M Li$_2$SO$_4$, 20% w/v PEG 3,000 (Figure II.3.c). Diffraction quality AidB crystals were optimized using the hanging drop method by mixing 2 μL protein with 1 μL precipitant solution (0.1 M HEPES, pH 7.5, 20% v/v ethanol, and 0.2 M MgCl$_2$) over 0.5 mL of precipitant solution at room temperature. Trigonal crystals

with dimensions of about 0.3 mm x 0.2 mm x 0.1 mm appeared in 2-3 weeks. No cryo-protectant was used prior to cryo-cooling the crystals in liquid nitrogen as the AidB crystals were fragile and dissolved quickly in solutions that were slightly different from the mother liquor. Crystal qualities varied greatly from one crystal to the next, and many crystals were screened to find the two best crystals that were used to obtain native datasets. Data were collected to 3.9 Å and 2.8 Å resolutions for two different crystals at Argonne Advanced Photon Source on beamline 24ID-C using Q315 ADSC detector. Data were processed and scaled using Denzo and Scalepack (Table II.2).

Both Selenomethionine (SeMet) AidB crystals and heavy atom derivatives of AidB crystals were used in an attempt to obtain experimental phases. Only small SeMet AidB crystals appeared in the same condition that yielded good diffracting native crystals, and both micro-seeding and streak seeding were used to improve the SeMet AidB crystal size. Upon streak seeding in 0.1 M HEPES, pH 7.5, 19% v/v ethanol, 0.25 M $MgCl_2$, plate-shaped SeMet AidB crystals were obtained (Figure II.3.d). However, these plate-shaped SeMet crystals diffracted to ~8-9.5 Å at the face of the plate, and almost no diffraction on the side of the plate. For heavy atom derivatives, native crystals (grown in 0.1 M tris, pH 7.5, 2% v/v tacsimate, 25% w/v PEG 3,000) were soaked in 1 mM of the following heavy atom solutions overnight: K_2PtCl_6, $KAu(CN)_2$, K_2HgI_4, $Sm(O_2C_2H_3)_3$, $Sm(NO_3)_3$, H_2PtCl_4, $Pb(NO_3)_2$ and $GdCl_3$. While crystals tolerated the heavy atom soaks without cracks or obvious crystal deterioration, none of these heavy atom soaked crystals diffracted.

Structure Determination and Refinement

The structure of AidB was solved by molecular replacement using PHASER[11] with the 2.8 Å resolution native dataset. A monomer of the tetrameric structure of AidB (3DJL; 100% identity), the same protein from *E. coli* published by T. Bowels *et al.*,[3] without any cofactors or waters was used as a search model. The best rotational and translational solution for the $P3_2$ space group had a correlation coefficient of 26.9 with 12 AidB molecules per asymmetric unit (3 tetramer units related by non-crystallographic symmetry). The PHASER solution was used in a rigid body refinement followed by simulated annealing refinement.[12] After the first round of refinement, the R and R_{free} were 31.9% and 31.5%, respectively. Water molecules were added to the model at 2.8 Å resolution followed by iterative rounds of model building and refinement using COOT and CNS, respectively.[12, 13] Non-crystallographic symmetry (NCS) restraints were used for each tetramer with high weight for the protein backbone and low weight for the bfactors, 300 and 2, respectively. Omit maps calculated using CNS simulated annealing were used to validate the model.[12] Refinement statistics for the final model is shown in Table II.3.

Final Protein Model

The final structure was refined to 2.8 Å resolution with R_{cryst} 23.9% and R_{free} 27.5%. Residues 2 – 540 were observed out of a total of 541 residues. There is one disulfide bond within each chain, between Cys28 and Cys540. Ramachandrain plot shows 88.5% of the residues are in the most-favored region, 10.6% of the residues in the allowed region, 0.5% of the residues in the generously allowed region, and 0.4% of the residues in the disallowed region.

Purification of Apo-AidB (FAD-free AidB)

A few methods were used to attempt to remove the FAD from AidB: (1) washing with 2 M KBr under slight denaturing condition, (2) dialysis under slight denaturing condition, (3) rechromatography, and (4) addition of $CaCl_2$. The purified FAD-containing AidB were first obtained as described in purification section above, further purification followed to remove FAD from the purified protein.

In the first attempt, FAD-containing protein was washed with KBr. The purified FAD-containing AidB was loaded onto a phenyl sepharose column (EMD Bioscience). While the protein was bound to the column, 10 column volumes (CV) of protein storage buffer (50 mM Tris, pH 7.8, 1 mM ethylenediaminetetraacetic acid (EDTA), 300 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol (BME)) with 2 M KBr and 2 M urea was used to wash the column. The protein bound on the column was then washed with 10 CV of storage buffer to remove the KBr and urea. Finally the protein was eluted in storage buffer with 0.4% w/v deoxycholate, and dialysis against storage buffer overnight was used to remove the deoxycholate.

In the second method, the purified FAD-containing AidB was dialyzed against storage buffer with 2 M urea. The dialysis buffer was changed twice over a 24 h period. The urea was removed from the protein by further dialysis against only the storage buffer.

In the third method, the purified FAD-containing AidB was rerun on the size exclusion column further in an attempt to remove the FAD. The Superose 6 (EMD Bioscience) was pre-equilibrated with 2 CV of AidB storage buffer prior to sample loading. The purified FAD-containing AidB was loaded onto the column and eluted in 1.5 CV of storage buffer. AidB eluted off the column in a single peak.

In the fourth method, 2 M $CaCl_2$ was used to degrade the FAD. Two methods of $CaCl_2$ addition were tested: (1) solid $CaCl_2$ was added directly to the FAD-containing AidB to a final concentration of 2 M, and (2) 3 volumes of 3.4 M $CaCl_2$ stock was mixed with 2 volumes of purified FAD-containing AidB protein. In both of these trials, the sample was mixed constantly during $CaCl_2$ addition. The solutions were left mixing at 4°C overnight. The precipitated protein was centrifuged and separated from the supernatant. Several buffers were used to redissolve the protein precipitate: AidB storage buffer, 0.1 M sodium phosphate buffer, pH 7.5, and 0.1 M disodium pyrophosphate buffer, pH 7.5. Apo AidB prepared with this method was used to carry out analytical ultracentrifugation experiment.

Sedimentation Velocity Experiments

Sedimentation velocity experiments were done with a Beckman Coulter Optima XL-I analytical ultracentrifuge with a Beckman An60Ti rotor, and a Beckman XL-A Monochrometer in the MIT Biophysical Instrumentation Facility. Absorbance data were collected using a velocity sedimentation run at 10°C at 30,000 rpm. Protein samples (wild type AidB and T185V/S191R/R324D triple mutant of AidB) were purified as described above, and stored in 10 mM Tris, pH 7.8, 300 mM NaCl, 1 mM EDTA, 10% glycerol, and 5 mM BME. The apo AidB was prepared by 2.0 M $CaCl_2$ precipitation to remove the FAD, and the precipitated protein was redissolved in 0.1 M disodium pyrophosphate, pH 7.5. Protein samples (6.1 μM and 7.5 μM of wild type AidB; 6.3 μM and 17.8 μM of T185V/S191R/R324D triple mutant of AidB; and 3 μM apo AidB) were loaded into one side of the sample cell, and the storage buffer without protein was loaded into the other side of the sample cell. Sednterp[14] and SEDANAL[15] were used for the data analyses of the wild type AidB and the triple mutant AidB. Sednterp[14] was used to calculate

values based on the experimental conditions and protein sequence for the density, partial specific volume, and the mass extinction coefficient at 280 nm. SEDANAL[15] was used to analyze the absorbance data from the sedimentation velocity experiments, along with the values obtain in Sednterp. DCDT+[16] was used for the data analysis of the apo AidB.

II.D. Results

Structural Overview

We have determined the structure of the FAD-containing protein AidB to 2.8 Å resolution in a P3$_2$ space group. The AidB structure obtained here belongs to a different space group compared to the one obtained by Eichman *et al.* (I222 space group).[3] While the crystal packing is different, the overall structures are similar to one another, with an RMSD of 0.34 Å over 538 C$_\alpha$ atoms. As noted previously, AidB contains a similar overall structural fold as members of the acyl CoA dehydrogenase family.[3] An AidB monomer consists of three major domains: an α-helical N-terminal domain (residues 25-64 and 86-161), a β-sheet domain in the middle (residues 198-282), and another α-helical C-terminal domain (residues 287-535) (Figure II.4). These three domains are arranged so that the α-helical regions are on one side of the monomer and the β-sheet domain is on the other side of the monomer (Figure II.4). The first 20 amino acids at the N-terminus are extended in a loop, which interacts with the neighboring monomer. There is little difference between each monomer within the asymmetric unit of the P3$_2$ cell; the RMSDs between each of 12 monomers range from 0.27 – 0.37 Å over 539 C$_\alpha$ atoms.

Monomers A and B form a tight dimer (Figure II.5.a-c). The buried surface area between monomers A and B is extensive, 3,459 Å^2, indicating that the dimeric interactions between A and B is unlikely to be a crystallographic artifact. This dimer interaction is also observed in the AidB

structure obtained by Eichman et al.[3] There is one FAD per AidB monomer, and the FADs bind at the interface of the AB dimer in an extended conformation. The isoalloxazine ring of the FAD is sandwiched between the α-helical region and the β-sheet region within a monomer, and the adenine end of the FAD interacts with the neighboring monomer.

On the basis of sequence and structural alignment, AidB belongs to the acyl CoA dehydrogenase family. Structural alignment of AidB with isovaleryl CoA dehydrogenase (IVD; sharing 24% identity and 44% homology) and with medium chain CoA dehydrogenase (MCAD; 24% identity and 43% homology) shows that the AB dimers of these three structures share a similar core structure. AidB has an RMSD of 1.68 Å over 314 C_α atoms with IVD, and an RMSD of 1.70 Å over 303 C_α atoms with MCAD. Unlike enzymes in the acyl CoA dehydrogenase family, AidB has an extra α-helix at the N-terminus (Figure II.6). The MCAD is shorter at the N-terminus, and IVD has only a short 3_{10} helix in place of AidB's first α-helix. In addition, AidB has two β-strands, antiparallel to each other, between α-helix3 and 4. Both IVD and MCAD lack these β-strands; a sequence gap is in place of these two β-strands. The biggest difference between AidB and enzymes in this family is the extra α-helical region, $\alpha14 - \alpha18$, at the C-terminus.

Enzymes in the acyl CoA dehydrogenase family are often tetramers. Although the dimeric interaction seen between AidB monomers A and B is conserved, the tetrameric arrangement of AidB differs from other acyl CoA dehydrogenases. In the crystal, AidB is a tetramer, composed of two dimers (AB and CD dimers) (Figure II.7.a); the same tetramer is observed in the AidB structure obtained by Eichman et al.[3] Due to the presence of the extra α-helices at the C terminus, the tetramer arrangement in AidB is different from other enzymes in the acyl CoA dehydrogenase family (Figure II.7.b). In AidB, the dimers are packed against each

other at the N-terminus; specifically, they are packed against α3, α4, β1 and β2. However, in IVD and MCAD, the dimers are packed against each other at the C-terminus; they are packed against α10-13.

FAD Binding Site

There are two FADs bound at the interface of monomers A and B. The FADs bind in the extended conformation; the isoalloxazine ring is buried between the α-helical half and the β-strand half of the monomer, and the adenine end of the FAD is at the interface of the α-helical halves of two monomers (Figure II.5.c). The same binding mode for FAD is observed in the acyl CoA dehydrogenase family. The FAD is somewhat solvent accessible (33% of the FAD is exposed to solvent); both at the isoalloxazine ring (~17%) and the adenine end (~16%) of the FAD (Figure II.8.b). The dimethylbenzene end of the isoalloxazine ring is buried as it in the acyl CoA dehydrogenase family.

The isoalloxazine ring of the FAD is between F216 and W424 (Figure II.9.b). Based on the structural alignment, both IVD and MCAD have an aromatic residue, W166, at the position corresponding to F216. However, W424 is not well conserved among these proteins; a Gly is in place of the Trp in IVD, and a Tyr is in place of the Trp in MCAD.[17] The Tyr in MCAD does not participate in aromatic stacking with the isoalloxazine ring of the FAD as the Trp in AidB does.[17]

The FAD-protein interactions are well conserved within the acyl CoA dehydrogenase family. There are few hydrogen bonding interactions to the FAD that are contributed by protein side chains, instead many are contributed by the backbone amides or carbonyls instead. The few hydrogen bond interactions contributed by the protein side chains are from Thr185, Ser191, and Ser218 from the same molecule, and Arg324 from the neighboring molecule (Figure II.9).

Thr185 hydrogen bonds with N1 of the isoalloxazine ring of FAD. This OH-N1 interaction is structurally conserved between IVD (Ser-OH in place of Thr-OH) and MCAD (Thr in the same position). Ser218 side chain hydrogen bonds with the N5 of the isoalloxazine ring of FAD (Figure II.9.a). This residue is a Thr in both IVD and MCAD, which is also able to contribute to hydrogen bonding to the FAD at N5 position. The hydroxyl of Ser191 hydrogen bonds with the ribityl O1 oxygen (Figure II.9.a). Arg324 side chain hydrogen bonds with the ribityl O2 oxygen (Figure II.9.a). Both Ser191 and Arg324 are strictly conserved between AidB, IVD and MCAD. Surprisingly, there are no hydrogen bond interactions between the protein and the adenine ring of the FAD; this part of the FAD is mostly stabilized by hydrophobic interactions. These interactions between AidB and the FAD are also observed by Eichman and coworkers in their structure.[3]

Both a photoreduction study on the FAD in AidB,[1] as well as an electrochemical study on the redox potential of the FAD in AidB (personal communication with Professor Sean Elliott's laboratory), show that the FAD can be reduced to a stable anionic semiquinone FAD species. From the structure, it appears that Thr185 could play two roles in stabilizing an anionic semiquinone species. The side chain hydroxyl group of Thr185 (2.8 Å away) is set up to donate a hydrogen bond to the N1 nitrogen of the isoalloxazine ring to stabilize a partial negative charge on the semiquinone FAD species. Furthermore, the amide backbone of Thr185, at 2.8 Å away from the C2 carbonyl oxygen of the isoalloxazine ring, could help further stabilize a partial negative charge in that oxygen through hydrogen bonding. Since the isoalloxazine ring binding site is surrounded by relatively hydrophobic residues with few positively charged residues to balance the negative charge on the anionic semiquinone, the solvent exposed nature of the FAD is likely to play a role in stabilizing the anionic semiquinone. Solvent exposure is known to

modulate redox potentials in flavodoxins.[18] Furthermore and most importantly, Ser218 is 2.87 Å away from the N5 nitrogen of the isoalloxazine ring (Figure II.9.a), and since serine is protonated at neutral pH, it may physically inhibit protonation at N5, leading to the anionic state of the semiquinone flavin. It is the protonation state of N5 that controls the charge of the flavin semiquinone.

Putative Substrate Binding Site

The CAVER program[19] was used to find a tunnel starting from the isoalloxazine ring of the FAD to the surface of the protein. A tunnel of approximately 25 Å long and 6.3-15 Å wide (the narrowest part of the tunnel is at the end near the W424) was found (Figure II.8.a). This tunnel could be the putative substrate binding site or an entryway for the putative substrate. The same tunnel is present in the IVD and the MCAD structures when they are superimposed onto the AidB structure (Figure II.8.c). In the MCAD structure, the substrate (octanoyl CoA) is bound in this site.[17] In the IVD structure, coenzyme A persulfide that co-purified with the enzyme is bound in this site.[20]

There is a cleft across the surface of the protein perpendicular to the tunnel (Figure II.8.b). The surface of the cleft is composed of many positively charged residues, such as Arg289, Lys293, Arg365, Arg375, Lys441, and Arg437. There are also positively charged residues on the inside surface of the tunnel, such as Lys293 and Arg299. As the tunnel goes deeper, the positive electrostatic nature of the tunnel changes to neutral. At the end of the tunnel is the negatively charged residue Glu425 (Figure II.8.a), which is in the same position as the catalytic base in MCAD based on structural alignment. In MCAD, the catalytic base (Glu376) initiates the proton

abstraction from the α-carbon in the substrate (Scheme I.1 in Chapter I). Since the substrate and the activity of AidB are unknown, it is unclear if Glu425 plays the same role in AidB.

A unique feature of the putative active site in AidB compared to IVD and MCAD is the number of cysteines it contains. There are three cysteines (Cys133, Cys 302 and Cys386) positioned below the isoalloxazine ring of the FAD (Figure II.9). None of these cysteines are present in the active site of IVD or MCAD.[17, 20] Cys133 is 5.51 Å away from the N5 nitrogen of the isoalloxazine ring. Cys302 is 8.90 Å away from Cys133, which is relatively far; but Trp424 is between these two cysteines, at 6.42 Å away from Cys302 and at 5.31 Å away from Cys133. Cys386 is a little further, 5.50 Å away from Cys302. This arrangement of cysteines so close to the isoalloxazine ring is intriguing but no function for these residues has been proposed.

Since AidB is on the adaptive response pathway for DNA repair, one hypothesis for the substrate of AidB is DNA. Hausinger *et al.* has shown by the use of gel shift assays that AidB binds to dsDNA,[1] and Eichman *et al.* proposed based on the structure and a site-directed mutagenesis study that dsDNA binds to the positively charged patch (consisting Arg461, Arg465, Arg468, Arg469, Arg475, and Arg518) at the C-terminus of the protein, the portion of the structure unique to AidB (Figure II.10.a).[3] This proposed DNA binding site is approximately 38 Å from the entrance of the tunnel where putative substrate binds (the distance is calculated by the shortest trajectory on the surface of the protein that goes from the middle of the positively charged patch of the proposed DNA binding domain to the entrance of the tunnel) (Figure II.10.b). Furthermore, this proposed DNA binding site is approximately 40 Å from the isoalloxazine ring of FAD (the distance is calculated by a straight line from middle of the positively charged patch to the isoalloxazine ring of FAD). Therefore, a damaged DNA strand binding at this proposed DNA binding site is too far from the putative substrate binding site for

the DNA lesion to be directly repaired. On the other hand, the positively charged cleft located outside the tunnel is more suitable as a DNA binding site for AidB if AidB is involved in direct DNA repair. A damaged DNA strand could bind at this positively charged cleft, and the lesion could enter the tunnel.

To examine whether a typical lesion is long enough to reach down this 25 Å long tunnel to the isoalloxazine ring of FAD for repair, possibly by dehydrogenase activity, an ethenoadenine lesion flipped out of dsDNA has been docked into this channel (Figure II.11). We find that a lesion flipped out of a dsDNA is not long enough to reach the FAD. Eichman *et al.* modeling results are consistent with our findings.[3] The shortest distance between the ethenoadenine and the N5 nitrogen of FAD is about 11.5 Å (Figure II.11). In order for this lesion to reach closer to the FAD, the dsDNA would have to be nicked and unwound. Alternatively, the substrate could be a damaged free nucleotide or a larger nucleotide adduct of DNA. Apart from DNA lesions as substrate for AidB, of course small molecules such as alkylating agents could also fit into the putative substrate binding channel.

AidB-ETF Docking Model

Assuming an ideal substrate for this channel exists, the next question is how the flavin is reoxidized. Once catalysis occurs in acyl CoA dehydrogenases (e.g., MCAD and IVD), an electron transfer flavoprotein (ETF) accepts the electrons from the reduced flavin of the acyl CoA dehydrogenase, and further transfers them to the respiratory chain.[21] Studies have shown that the electron transfer between acyl CoA dehydrogenases and ETF occurs by forming a transient complex while the product of the acyl CoA dehydrogenase is still bound to the protein.[22] To examine whether *E. coli* AidB might interact with an ETF, a docking model of

AidB with the ETF from human has been created. The human ETF is chosen for the docking for several reasons: (1) there is no *E. coli* ETF structure known; (2) the full structure of human ETF alone is known; and (3) the structure of human ETF complex with MCAD is known, although the FAD binding domain of the ETF in the MCAD-ETF complex is disordered (Figure II.12.a). The docking of AidB with the human ETF is generated first by superposition of the full length structure of the human ETF (1EFV) onto the MCAD-ETF complex (1T9G), to predict the location of the missing FAD-binding domain of ETF with respect to the MCAD-ETF complex (Figure II.12.b). Second, the AidB tetramer is superimposed using the AB dimer onto the orange dimer of the MCAD-ETF complex (Figure II.12.b). This superposition orients the AidB tetramer with respect to the ETF. Combining the two superpositions together, a docking model of AidB with the human ETF is generated (Figure II.12.c).

The first thing that is noticeable about the AidB-ETF tetrameric docking model is that the CD dimer of the AidB tetramer clashes with the FAD-binding domain of the ETF (Figure II.12.c). Due to the different tetramerization modes of MCAD and AidB, no clash is observed between the MCAD dimer colored in deep olive and ETF (Figure II.12.b). The FAD binding domain of ETF has two minor clashes with monomer B of AidB in the docking model (Figure II.13.a); one is between the C-terminal residues (β238-244) of the human ETF and the N-terminal residues (4-9) of monomer B of AidB, and another is between a turn (α303-306) of human ETF and the N-terminal residues (11-13) of monomer B of AidB. These minor clashes are also observed in the superposition of the FAD binding domain of ETF onto the MCAD-ETF complex, which might be due to the slightly different orientation of the FAD binding domain of ETF assumes when not bound to its partner protein.[21]

Excluding the minor clashes, there is high overall shape complementarity between the AidB dimer and human ETF (Figure II.13.a-c). There are only a few interactions observed between ETF and MCAD in the complex, which agrees with the kinetics data that ETF only forms a transient complex with its dehydrogenase partner.[22] The main interaction is between residues 191-200 of the β-subunit of the ETF, called the recognition loop, and a hydrophobic patch on its partner.[21] This interaction is satisfied in the AidB-ETF docking model (Figure II.13.b); the recognition loop of ETF fits in the hydrophobic patch on the monomer B of AidB. Both kinetics and structural studies have shown that once the recognition loop of the ETF binds to its partner, the FAD-binding domain of ETF is free to sample a wide range of conformational space to find a position that is suitable for transient electron transfer.[22] Since ETF binds many acyl CoA dehydrogenase partners, the interactions between the FAD-binding domain of ETF and its partner protein are promiscuous.[22]

The closest distance between the FAD of AidB monomer A (C6 carbon of the isoalloxazine ring) and the FAD of human ETF (N3 nitrogen of the adenine ring) is 21 Å. The shortest distance between the isoalloxazine ring of AidB monomer A and that of human ETF is 33 Å, whereas the shortest distance between the isoalloxazine ring of AidB monomer B and that of human ETF is 38 Å (Figure II.13.c). These distances are comparable to what one would expect to find in the MCAD-ETF complex, where the closest distance between the two isoalloxazine rings is predicted to be 35 Å.[21] Although a 33 Å – 35 Å distance is long for electron transfer, the ETF partner is expected to go through a conformational change that brings the two FADs closer together during electron transfer, an idea consistent with the ETF FAD domain being disordered in the MCAD-ETF crystal structure. While there are no data on

whether AidB binds to an ETF, this docking model suggests that such interaction or complex is possible, and further study should be focused on determining if AidB interacts with ETF.

Design and Oligomeric State Analysis of the T185V/S191R/R324D Triple Mutant of AidB

Since the FAD binds at the interface of two monomers (A and B), an interesting question that arises is whether FAD binding is required for AidB dimerization. The answer to this question will also provide insight into whether the role of FAD is at least partially structural in nature. To address this issue, a triple mutant was designed to knock out FAD binding, and the oligomeric state of this mutant protein was analyzed by analytical ultracentrifugation.

Since there are only a few hydrogen bonding interactions between the protein side chain and the FAD, with most of the interactions provided by the backbone atoms, designing a mutant that no longer binds FAD was difficult. Thr185 contribute two hydrogen bonds to the FAD through its side chain hydroxyl group (Figure II.9.a), mutating Thr185 to a Val would eliminate the two hydrogen bonds, and the structure of Val is sterically the same as Thr, and thus this change not expected to perturb the overall protein structure.

The hydroxyl group of Ser191 hydrogen bonds with one of the oxygen atoms on the phosphate group of FAD (Figure II.9.a). Ser is a relatively small amino acid, thus allowing room for the phosphate of FAD to be present. However, if Ser191 is mutated to a large side chain amino acid, such as Arg, the side chain of Arg is expected to extend to the same space that is occupied by the FAD; but yet not large enough to extend into the space of the neighboring AidB molecule to prevent dimerization (assuming dimerization does not require the presence of FAD). Since many of the other hydrogen bond interactions to FAD contributed by the protein backbone are still present in the mutant protein, it may be important to have a final mutation that sterically

blocks FAD binding rather than just destabilizing FAD binding. Therefore, the last mutation involves changing Arg324 to an Asp. The positively charged Arg side chain hydrogen bonds with the negatively charged phosphate oxygen on FAD (Figure II.9.a). Asp, on the other hand, is a negatively charged residue at neutral pH, and thus should repel the negatively charge phosphate oxygen on FAD. Additionally, the side chain of Asp is smaller than Arg, and should not perturb the overall AidB structure.

The T185V/S191R/R324D triple mutant of AidB was created by site-directed mutagenesis, and purified using the same method as used for the wild type AidB. The mutations have successfully eliminated FAD binding, based on the clear color of the solution of the final purified triple mutant protein, suggesting the absence of FAD.

An analytical ultracentrifugation experiment was used to determine if the T185V/S191R/R324D triple mutant of AidB is a monomer or dimer in solution. From a sedimentation velocity experiment for the triple mutant of AidB at concentrations of 6.3 µM and 17.8 µM, the best fit sedimentation coefficient (S) is 6.9, and the best fit molecular weight is 63±1.1 kDa (Figure II.14). The best fit molecular weight of the triple mutant of AidB is similar to the theoretical molecular weight of AidB, 61 kDa, suggesting that the triple mutant of AidB is a monomer in solution at 6.3 µM and 17.8 µM. These data imply that in the absence of FAD binding, AidB is a monomer. Ideally, further biophysical analyses on this mutant protein are needed to prove that the mutations affect only the FAD binding ability and not the structure of the AidB monomer. Such biophysical analyses could include circular dichroism (CD) to determine the overall proper folding of the mutant protien, crystallization to determine structural impact of the mutations, and isothermal titration calorimetry (ITC) to measure the FAD binding constant.

Purification and Analytical Ultracentrifugation Analysis of Apo-AidB (FAD-free AidB)

Since we were unsure whether the mutations perturb the protein structure, measuring the oligomeric state of apo (FAD-free) AidB would also address the issue of whether FAD binding is required for dimerization. Four different methods were tried to remove the FAD from AidB. In the first method, 2 M KBr and 2 M urea were used to wash the FAD-containing protein while it was bound to a phenyl sepharose column. About half of the protein came off the column during the KBr/urea wash, probably because of the slightly denaturing conditions. The protein that washed off the column early still had its FAD bound to the protein, judging by the yellow color of the protein solution. The other half of the protein stayed bound to the column during the KBr/urea wash. However, this protein also had FAD bound when it was finally eluted from the column, again judging by the yellow color of the protein solution. The protein that stayed bound to the column was most likely properly folded, such that the FAD could not be easily removed.

Dialysis sometimes can remove FAD from FAD-containing proteins if the flavin is loosely bound.[23] Since the dialysis membrane only allows small molecules like FAD or ions to pass through, but not large molecules like protein, loosely bound FAD can be removed from the protein in this way. When the FAD-containing AidB was dialyzed against a large volume (4 L) of storage buffer overnight, with two buffer changes in between, the FAD remained bound to the protein after 24 hrs of dialysis, judging by the yellow color of the protein solution. Some studies have shown that extensive dialysis (for as long as a month) with frequent buffer changes are sometimes necessary to remove the FAD from the protein.[23] Extensive dialysis was not tried, and should be reexamined in the future if other methods fail to remove the FAD from AidB.

Another method to remove FAD from other FAD-containing proteins is the use of multiple chromatography rounds. When AidB was purified twice using the size exclusion

chromatography, the protein solution still had a faint yellow color, indicating that FAD was still bound to the protein.

Lastly, adding $CaCl_2$ to protein has been shown to successfully remove FAD from xanthine oxidase[24]. High concentration of $CaCl_2$ is believed to promote hydrolysis of FAD, resulting in a degraded form of FAD that does not bind to proteins.[24] One drawback of this method is that $CaCl_2$ is known to cause copious protein precipitation.[24] Although protein precipitates can be redissolved in various buffers, the yield of FAD-free protein is low. Two ways of adding $CaCl_2$ to AidB were tested to minimize protein precipitation; (1) solid $CaCl_2$ was added directly to the FAD-containing AidB to a final concentration of 2 M, and (2) 3 volumes of 3.4 M $CaCl_2$ stock was mixed with 2 volumes of purified FAD-containing AidB protein. The mixing of 3 parts $CaCl_2$ stock with 2 parts AidB protein seemed to better minimize protein precipitation as judged by the size of the precipitated protein pellet. Of the various buffers (AidB storage buffer; 0.1 M sodium phosphate buffer, pH 7.5; and 0.1 M sodium pyrophosphate buffer, pH 7.5) that were tested to redissolve the AidB precipitate, sodium pyrophosphate worked the best. This method was successful in removing the FAD from AidB, as judged by the lack of yellow color of the protein solution. However, this method yielded little apo AidB for biophysical experiments. Using this method without optimization, an 8 L protein preparation yielded 1 mL of 3.2 μM of apo AidB, just enough apo protein to do one analytical ultracentrifugation experiment. Further optimization, or doing large protein preparations this way, will be required to obtain enough apo AidB for more biophysical experiments, such as analytical ultracentrifugation with different protein concentrations to determine the oligomeric state of apo AidB and isothermal titration calorimetry (ITC) to measure the FAD binding constant.

The sedimentation velocity experiments for apo AidB at a concentration of 3 μM. Yielded a best fit sedimentation coefficient (S) of 6.7, and the best fit molecular weight of 66±1.0 kDa (Figure II.15). The best fit molecular weight of the apo AidB is similar to the theoretical molecular weight of AidB, 61 kDa, suggesting that the apo AidB is a monomer in solution at 3 μM. Although these data are consistent with the triple mutant data indicating that AidB is a monomer in the absence of FAD binding, these data are preliminary and this experiment must be repeated.

Oligomeric State of Wild Type AidB in the Crystal

The AidB structure presented here has a different crystal packing compared to the one solved by Eichman and coworkers.[3] This AidB structure belongs to the P3$_2$ space group, and the asymmetric unit contains 12 monomers of AidB (3 tetramers related by a 3-fold non-crystallographic symmetry) (Figure II.16). The interfaces between monomers of the same color are also observed in a different crystal form, I222, observed by Eichman and coworkers.[3] However, the interfaces between monomers of different colors (e.g., between a blue and a green molecule, or a green and a pink molecule, or a blue and a pink molecule) have not been observed previously. In the I222 space group that Eichman and coworkers obtained, AidB can form higher order oligomers; four AidB tetramers form a 25 Å pore with the proposed DNA binding region of two tetramers on the side of the pore (Figure II.1).[3] The "mega complex" of AidB, generated by crystal packing, led to a proposed model that DNA can be sequestered in these pores, thus shielding the DNA from alkylating agents in the environment.[3] This model seems highly unlikely for several reasons that will be discussed in more detail later. But based on the crystal packing observed here, AidB tetramers do not form the same higher order oligomers as the one Eichman

observed in I222 crystals, making it more unlikely that the crystal packing Eichman observed is physiologically relevant.

In the AidB tetramer observed here, the buried surface area between monomers A and B is extensive, 3,459 Å^2 (Figure II.17.a and b), consistent with a physiologically relevant dimer. A large number of interactions between these monomers are hydrogen bonding interactions, listed in Table II.4. On the other hand, the buried surface area between monomers A and C or B and D is only 883 Å^2. Contrary to the AB dimer (32 putative hydrogen bonds), there are only three putative hydrogen bonds between monomers A and C, listed in Table II.5. There seems to be little interactions stabilizing the AC or BD dimer, thus raises the question of whether the tetramer is a relevant oligomeric state for AidB in solution.

Analytical Ultracentrifugation of Wild Type AidB

To identify the relevancy of the AC or BD dimer observed in the crystal, analytical ultracentrifugation characterization was used to determine the oligomeric state of wild type FAD-bound AidB in solution. From the sedimentation velocity experiments for wild type AidB at 6.1 μM and 7.5 μM, the best fit sedimentation coefficient (S) is 5.0, and the best fit molecular weight is 122±1.3 kDa (Figure II.18). The theoretical molecular weight of AidB monomer is 61 kDa, about half of the experimentally fit molecular weight. These data imply that wild type AidB is a dimer in solution at concentrations of 6.1 μM and 7.5 μM.

Our analytical ultracentrifugation data suggest that the wild type AidB is a dimer, while both crystal structures show a tetramer; other literature reports say that AidB is a tetramer based on analytical ultracentrifugation or gel filtration, but data are not shown.[1, 3] The concentration of AidB present in the analytical ultracentrifugation and crystallographic experiments might

account for the discrepancy in these results. In our analytical ultracentrifugation experiments, less than 1 mg/mL of protein was present. Under such low concentration, the interactions between the AB and CD dimer discussed above may not be strong enough to keep the protein in a tetrameric state. On the other hand, 10 mg/mL of protein solution was used for crystallization, and the concentration of AidB would most likely be greater than 10 mg/mL in the crystal. Under such high concentration, AidB forms a tetramer at least in the crystalline state. Whether the oligomeric state of AidB is a dimer (as observed in our analytical ultracentrifugation experiment) or tetramer (as observed in the crystal structures and as previously reported),[1, 3] the oligomeric state under physiological condition is likely to depend on the local concentration of AidB in the cell.

II.E. Discussion

AidB is an adaptive response flavoprotein that was discovered several decades ago.[10] Although AidB has been shown to be up-regulated in the presence of alkylating agents, its function is still unknown beyond its ability to bind DNA non-specifically. We solved the crystal structure of AidB in a different space group ($P3_2$) from the one reported previously (I222) by Eichman et al.[3] Structural analysis of AidB, along with some biochemical work, has has allowed us to reconsider the several possible functions for AidB. Proteins usually achieve their function non-enzymatically, as in the case of transcription factors, or enzymatically, as in the case of dehydrogenases, or both, as in the case of the Ada protein. One hypothesis is that AidB could prevent DNA damage by sequestering DNA from alkylating agents, as in the model proposed by Eichman et al.[3] Another possibility would be that AidB acts as an enzyme to deactivate alkylating agents or repair alkylated DNA damage lesions. Or, AidB could function like the Ada

protein, where it is an enzyme that repairs DNA methylation damage and then becomes a transcription factor upon transfer of a methyl group to its active site cysteine.[25]

The AidB crystal structure solved here contains the same dimeric and tetrameric arrangement of AidB as the ones observed by Eichman *et al.*; however, the tetrameric units do not form a pore with DNA binding domains pointing toward each other as Eichman proposed (Figure II.1). In the crystal packing observed here, three tetramer units of AidB are packed with a central non-crystallographic three-fold axis (Figure II.16). It is energetically costly for AidB to protect only a small section of the dsDNA from alkylating agents by secluding the DNA from the environment using four tetramer units of AidB (16 monomers). Additionally, from our biochemical study, it is unclear if AidB will be a tetramer under cellular concentrations. AidB does form a tetramer when present in high concentrations (about 200 μM) as in crystallization. On the other hand, analytical ultracentrifugation experiments show AidB is a dimer in lower concentrations (about 7 μM). These biochemical experiments suggest the oligomeric state of AidB might depend on its concentration. The concentration of AidB in the *E. coli* cell before and after an adaptive response is not known. However, it is known that the level of Ada protein is about 3000 molecules per cell after the adaptive response,[26] and that Ada is up-regulated more than AidB.[27] Given these findings, 3000 molecules is an upper estimate for the number of molecules of AidB in *E. coli* under stress due to DNA damaging agents. Assuming a 10^{-15} L volume for an *E. coli* cell, the upper estimate concentration for AidB is 5 μM, a concentration consistent with a dimeric state. Since the concentration of AidB inside the cell has not been measured, it would be premature to conclude that AidB could form the "mega complex" that Eichman *et al.* proposes. Furthermore, the desolvation energy for removing the water molecules around the DNA prior to binding in the pore is energetically high; raising questions about the

physiological relevance of a DNA sequestration model. The AidB "mega complex" in the previously observed space group (I222) is most likely a coincidental result of crystallographic packing.

The AidB structure contains several characteristics that suggest that the protein could be an enzyme that carries out a reaction similar to that catalyzed by acyl CoA dehydrogenases. First, the AidB dimer structure shares a similar structural fold and high sequence homology with other enzyme dimers in the acyl CoA dehydrogenase family; however, the tetramer unit is assembled differently. Two AidB dimers are packed against their N-terminus, whereas the acyl CoA dehydrogenase dimers, such as IVD and MCAD, are packed against their C-termini. Secondly, AidB binds one FAD per monomer, just as the acyl CoA dehydrogenases do. In both cases, the FAD binds in an extended conformation at the dimer interface with the dimethylbenzene side of the flavin ring buried from solvent. Lastly, based on structural alignment with other acyl CoA dehydrogenases, AidB also contains the conserved glutamate residue (Glu425), which is the catalytic base in other acyl CoA dehydrogenases that initiates proton abstraction from the α-carbon of the substrate. These structural similarities suggest that AidB could function as an enzyme with acyl CoA dehydrogenase-like activity. In fact, AidB does exhibit low levels of IVD activity.[2]

Since the function of the AidB protein is unknown, it is unclear if the role of the FAD is structural or catalytic. As discussed above, the structural relationship with acyl CoA dehydrogenases suggests that the role of FAD is catalytic, but analytical ultracentrifugation results discussed below suggest that it could play a structural role as well. In terms of the AidB FAD, both a photoreduction study[1] and an electrochemical study (personal communication with Professor Sean Elliott's laboratory), show that the FAD is redox-active, and that it can be

reduced to a stable anionic semiquinone FAD species. From the structure, residues Thr185 and Ser218 are likely to play a role in stabilizing an anionic semiquinone species. In particular, Ser218 may block protonation of N5 by acting as a hydrogen bond donor to the deprotonated N5 of isoalloxazine ring. These important hydrogen bonding contacts are conserved between AidB and the acyl CoA dehydrogenase family.

If the role of FAD in AidB is catalytic, then AidB must have a way to reoxidize either the anionic semiquinone or $FADH_2$ to regenerate the FAD for another round of catalysis. For other members of the acyl CoA dehydrogenase family, electrons are transferred one at a time to an electron transfer flavoprotein (ETF). Kinetic studies have shown that electron transfer between ETF and acyl CoA dehydrogenases occurs when product is still bound through a transient complex formed by these proteins.[28, 29] One ETF usually interacts with several acyl CoA dehydrogenases. For example, the human ETF interacts with nine acyl CoA dehydrogenases, including IVD and MCAD.[30] Since there is no crystal structure of ETF from *E. coli*, and the complex structure of human ETF with MCAD is known, the human ETF was used to generate a docking model with AidB.

From the AidB-ETF docking model, it is obvious that the CD dimer interferes with ETF binding to the AB dimer of AidB. However, ETF binding does not interfere with MCAD tetramerization due to the difference in its tetrameric arrangement. If ETF interacts with AidB, it would interact with the AidB dimer. The docking model with an AidB dimer and ETF shows high overall shape complementarity. There is only a minor clash between a few residues at the N-terminus of AidB and the FAD binding domain of ETF (Figure II.13.a). Such minor clashes are also present when the FAD binding domain of ETF is superimposed onto the structure of the MCAD-ETF complex where the FAD binding domain of ETF is disordered. The interaction

between the "recognition loop" of ETF (β191-β200) and a hydrophobic patch on MCAD is conserved in the AidB-ETF docked model (Figure II.13.c). The fact that the structure of AidB docks so well with the structure of an ETF, taken together with the similarities in the FAD-protein interactions, suggests that the re-oxidation process used by acyl Co dehydrogenases may also be in play here.

The structure of AidB suggests a putative substrate binding tunnel similar to the one found in acyl CoA dehydrogenases. The tunnel is approximately 25 Å long and 6.5-15 Å wide. The DNA binding ability of AidB and its adaptive response to alkylating agents suggest one possible substrate for AidB would be damaged DNA. An ethenoadenine base flipped out of dsDNA was docked in the putative substrate binding tunnel, and it was found that a flipped base is too short to reach the FAD site. Alternatively, the substrate could be a damaged free nucleotide or a larger nucleotide adduct of DNA. The tunnel is relatively positive charged, and thus a negatively charged DNA substrate is a reasonable possibility. Small molecule substrates, such as DNA damaging agents, would also fit inside the putative substrate tunnel. However, the tunnel seems to be too large for one small molecule to block off the active site from the outside environment during catalysis.

The entrance of the tunnel that leads to the putative active site is 38 Å away from the DNA binding domain proposed by Eichman *et al.*, suggesting that AidB could be a dual-function protein. AidB could be like the Ada protein, which has only one catalytic turnover to demethylate DNA lesions by directly transferring the methyl group to an active site cysteine residue, at which point it turns into a transcription factor and binds to DNA for gene regulation. In comparison, AidB also has cysteines (Cys133, Cys302 and Cys 386) near the putative active

site, especially Cys133 at 5.5 Å away from N5 nitrogen of the isoalloxazine ring seems intriguing.

Although the electrochemistry and the structural environment of the AidB FAD suggest that its role is likely to be catalytic, that does not rule out a second role for FAD in supporting dimerization. Based on analytical ultracentrifugation experiments, the triple mutant of AidB that is deficient in FAD binding is also monomeric in solution, whereas the holo- (FAD-containing) wild type AidB is dimeric. These data suggest that FAD induces AidB dimerization. However, FAD does not seem to have an influence on dimerization of other acyl CoA dehydrogenases. For example, the apo butyryl-CoA dehydrogenase exists as a dimer in solution, and it reassociates as a tetramer in the presence of excess FAD.[31] On the other hand, the short-chain acyl CoA dehydrogenase (SCAD), MCAD, and long-chain acyl CoA dehydrogenase (LCAD) have all been shown to be tetramers in the apo form.[32] An improved method for obtaining apo (FAD-free) AidB will be necessary to fully address the influence of FAD binding on the AidB oligomeric state.

Determining the function of AidB has been challenging due to the lack of reproducible phenotypes *in vivo*.[33] While there is a report that knocking out AidB leads to higher sensitivity to MNNG,[2, 10] this finding could not be repeated, even when using the same knock-out cells.[33] In the absence of a phenotype, the structure of AidB has allowed us to reconsider what the function of AidB might be, although it is premature to settle on any one of them. Further studies will be required to test these possibilities. For example, computational experiments on the energetics of DNA desolvation would be useful in evaluating the likelihood of DNA being sequestered by a mega protein complex like the one proposed by Eichman *et al.* Biochemical studies should be conducted with different lesions, whose toxicity or mutagenicity could be alleviated by

dehydrogenation-like chemistry, as potential AidB substrates. Also, determining if AidB interacts with an *E. coli* ETF could provide insight as to whether the function of AidB is more closely related to dehydrogenases or oxidases; such studies should be done under concentrations where AidB exists as a dimer. As for thinking about the role of FAD, other studies that might further dissect the influence of FAD binding on AidB dimerization would be useful for considering the possibility of a dual-function protein.

Every cell encounters frequent endogenous and environmental carcinogens. AidB is up-regulated, along with other DNA repair enzymes (AlkA, AlkB and Ada), in the presence of alkylating agents, suggesting that AidB plays an important role in the cellular response to DNA damage. Understanding of the role of each player in maintaining the genomic integrity is important for preventing genetic diseases, cancers, and cell deaths that might arise from damaged DNA.

II.F. Acknowledgements

II.G. References

1. Rohankhedkar, M. S.; Mulrooney, S. B.; Wedemeyer, W. J.; Hausinger, R. P., The AidB component of the Escherichia coli adaptive response to alkylating agents is a flavin-containing, DNA-binding protein. *Journal of Bacteriology* **2006**, 188, (1), 223-230.

2. Landini, P.; Hajec, L. I.; Volkert, M. R., Structure and Transcriptional Regulation of the Escherichia-Coli Adaptive Response Gene Aidb. *Journal of Bacteriology* **1994**, 176, (21), 6583-6589.

3. Bowles, T.; Metz, A. H.; O'Quin, J.; Wawrzak, Z.; Eichman, B. F., Structure and DNA binding of alkylation response protein AidB. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, (40), 15299-15304.

4. Sedgwick, B., Repairing DNA-methylation damage. *Nature Reviews Molecular Cell Biology* **2004**, 5, (2), 148-157.

5. O'Brien, P. J.; Ellenberger, T., The Escherichia coli 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *Journal of Biological Chemistry* **2004**, 279, (26), 26876-26884.

6. Falnes, P. O., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Research* **2004**, 32, (21), 6260-6267.

7. Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B., Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002**, 419, (6903), 174-178.

8. Delaney, J. C.; Smeester, L.; Wong, C. Y.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M., AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nature Structural & Molecular Biology* **2005**, 12, (10), 855-860.

9. Frick, L. E.; Delaney, J. C.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Alleviation of 1,N-6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, (3), 755-760.

10. Volkert, M. R.; Nguyen, D. C., Induction of Specific Escherichia-Coli Genes by Sublethal Treatments with Alkylating-Agents. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1984**, 81, (13), 4110-4114.

11. Read, R. J., Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D-Biological Crystallography* **2001**, 57, 1373-1382.

12. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson,

T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **1998,** 54, 905-921.

13. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **2004,** 60, 2126-2132.

14. Laue, T. M.; Stafford, W. F., Modern applications of analytical ultracentrifugation. *Annual Review of Biophysics and Biomolecular Structure* **1999,** 28, 75-100.

15. Stafford, W. F.; Sherwood, P. J., Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophysical Chemistry* **2004,** 108, (1-3), 231-243.

16. Philo, J. S., A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Analytical Biochemistry* **2000,** 279, (2), 151-163.

17. Kim, J. J. P.; Wang, M.; Paschke, R., Crystal-Structures of Medium-Chain Acyl-Coa Dehydrogenase from Pig-Liver Mitochondria with and without Substrate. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** 90, (16), 7523-7527.

18. Swenson, R. P.; Krey, G. D., Site-Directed Mutagenesis of Tyrosine-98 in the Flavodoxin from Desulfovibrio-Vulgaris (Hildenborough) - Regulation of Oxidation-Reduction Properties of the Bound Fmn Cofactor by Aromatic, Solvent, and Electrostatic Interactions. *Biochemistry* **1994,** 33, (28), 8505-8514.

19. Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J., CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *Bmc Bioinformatics* **2006,** 7, -.

20. Tiffany, K. A.; Roberts, D. L.; Wang, M.; Paschke, R.; Mohsen, A. W. A.; Vockley, J.; Kim, J. J. P., Structure of human isovaleryl-CoA dehydrogenase at 2.6 angstrom resolution: Basis for substrate specificity. *Biochemistry* **1997,** 36, (28), 8455-8464.

21. Toogood, H. S.; van Thiel, A.; Basran, J.; Sutcliffe, M. J.; Scrutton, N. S.; Leys, D., Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-CoA dehydrogenase complex. *J Biol Chem* **2004,** 279, (31), 32904-12.

22. Toogood, H. S.; Leys, D.; Scrutton, N. S., Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. *Febs J* **2007,** 274, (21), 5481-504.

23. Malhotra, K.; Kim, S. T.; Walsh, C.; Sancar, A., Roles of Fad and 8-Hydroxy-5-Deazaflavin Chromophores in Photoreactivation by Anacystis-Nidulans DNA Photolyase. *Journal of Biological Chemistry* **1992,** 267, (22), 15406-15411.

24. Komai, H.; Massey, V.; Palmer, G., The preparation and properties of deflavo xanthine oxidase. *J Biol Chem* **1969,** 244, (7), 1692-700.

25. Sedgwick, B.; Lindahl, T., Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* **2002,** 21, (58), 8886-8894.

26. Nieminuszczy, J.; Grzesiuk, E., Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA. *Acta Biochimica Polonica* **2007,** 54, (3), 459-468.

27. Landini, P.; Busby, S. J., Expression of the Escherichia coli ada regulon in stationary phase: evidence for rpoS-dependent negative regulation of alkA transcription. *J Bacteriol* **1999,** 181, (21), 6836-9.

28. Chohan, K. K.; Jones, M.; Grossmann, J. G.; Frerman, F. E.; Scrutton, N. S.; Sutcliffe, M. J., Protein dynamics enhance electronic coupling in electron transfer complexes. *Journal of Biological Chemistry* **2001,** 276, (36), 34142-34147.

29. Toogood, H. S.; van Thiel, A.; Scrutton, N. S.; Leys, D., Stabilization of non-productive conformations underpins rapid electron transfer to electron-transferring flavoprotein. *Journal of Biological Chemistry* **2005,** 280, (34), 30361-30366.

30. Roberts, D. L.; Frerman, F. E.; Kim, J. J. P., Three-dimensional structure of human electron transfer flavoprotein to 2.1-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** 93, (25), 14355-14360.

31. Van Berkel, W. J.; Van den Berg, W. A.; Muller, F., Large-scale preparation and reconstitution of apo-flavoproteins with special reference to butyryl-CoA dehydrogenase from Megasphaera elsdenii. Hydrophobic-interaction chromatography. *Eur J Biochem* **1988,** 178, (1), 197-207.

32. Ikeda, Y.; Okamura-Ikeda, K.; Tanaka, K., Purification and characterization of short-chain, medium-chain, and long-chain acyl-CoA dehydrogenases from rat liver mitochondria. Isolation of the holo- and apoenzymes and conversion of the apoenzyme to the holoenzyme. *J Biol Chem* **1985,** 260, (2), 1311-25.

33. Frick, L. E.; Massachusetts Institute of Technology. Biological Engineering Division. The versatile E. coli adaptive response protein AlkB mitigates toxicity and mutagenicity of etheno-, ethano-, and methyl-modified bases in vivo. Thesis Ph. D. --Massachusetts Institute of Technology Biological Engineering Division 2007., 2007.

II.H. Tables, Schemes and Figures

Table II.1. Primer pairs for each AidB mutants.

AidB Mutants	Primer pairs
T185V/S191R/R324D Triple mutant	T185V-S191R-F: 5'-GGCATGGGAATGGTGGAAAAGCAGGGCGGTAGGGATGTTATGAGC-3' T185V-S191R-R: 5'-GCTCATAACATCCCTACCGCCCTGCTTTTCCACCATTCCCATGCC-3' R324D-F: 5'-TATCATGCACATCAAGACCATGTTTTTGGTAAT-3' R324D-R: 5'-ATTACCAAAAACATGGTCTTGATGTGCATGATA-3'
T185V/S191R/R324D/W424A Quadtriple mutant	T185V-S191R-F: 5'-GGCATGGGAATGGTGGAAAAGCAGGGCGGTAGGGATGTTATGAGC-3' T185V-S191R-R: 5'-GCTCATAACATCCCTACCGCCCTGCTTTTCCACCATTCCCATGCC-3' R324D-F: 5'-TATCATGCACATCAAGACCATGTTTTTGGTAAT-3' R324D-R: 5'-ATTACCAAAAACATGGTCTTGATGTGCATGATA-3' W424A-F: 5'-CCGGTAAACAGTATTGCGGAAGGTTCCGGCAAT-3' W424A-R: 5'-ATTGCCGGAACCTTCCGCAATACTGTTTACCGG-3'
C133S mutant	C133S-F: 5'-GAGGCAGGGTCGTTATCTCCGATAACCATGACC-3' C133S-R: 5'-GGTCATGGTTATCGGAGATAACGACCCTGCCTC-3'
C302S mutant	C302S-F: 5'-ATGACGCGTTTTGATTCCGCCCTGGGTAGCCAT-3' C302S-R: 5'-ATGGCTACCCAGGGCGGAATCAAAACGCGTCAT-3'

Table II.2. Data collection and processing statistics for two AidB native datasets.

Structure	AidB Native	AidB Native2
Crystal Parameter		
Space Group	$P3_2$	$P3_2$
Unit cell dimension (Å)	179.72, 179.72, 204.24 90, 90, 120	179.62, 179.62, 209.10 90, 90, 120
# of molecules per ASU	12	12
Data Collection		
Wavelength (Å)	1.0	1.0
Resolution limit[b] (Å)	50.0 – 2.80 (2.90 – 2.80)	50.0 – 3.90 (4.04 – 3.90)
R_{sym}[a,b] (%)	9.5 (45.7)	9.7 (30.7)
No. of unique reflections	174735	65147
Redundancy[b]	6.1 (5.1)	9.2 (4.9)
Completeness[b] (%)	96.2 (97.4)	97.1 (75.1)
I/sigma[b]	14.1 (3.3)	24.5 (5.2)

[a]$R_{sym} = (\Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|)/ \Sigma_{hkl}\Sigma_i I_i(hkl)$ for n independent reflections and i observations of a given reflection. $<I(hkl)>$ = average intensity of the ith observation.

[b]Numbers for the highest resolution shell are shown in parentheses.

Table II.3. Model refinement statistics for AidB structure with the native dataset.

Resolution range	50.0 – 2.80
$R_{\text{cryst}}/R_{\text{free}}{}^{\text{a}}$ (%)	23.9/27.5
No. of protein atoms	50838
No. of FAD molecules	12
No. of water molecules	773
Rmsd bond lengths (Å)	0.0076
Rmsd bond angles (°)	1.42
Ramachandran analysis	
Most favored (%)	88.5
Allowed (%)	10.6
Generously allowed (%)	0.5
Disallowed (%)	0.4
B-factors (Å^2)	
Protein	43
FAD	50
Water	35

$^{\text{a}}R_{\text{cryst}} = \Sigma_h||F_o(h)| - |F_c(h)|| \,/\, \Sigma_h|F_o(h)|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is calculated the same way with a test set of reflections (10%) that are not used during refinement.

Table II.4. List of putative hydrogen bond distances between molecule A and B of AidB. These distances are calculated in PISA Server, http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

	Monomer A residue	Distance (Å)	Monomer B residue
1	ARG 324[NH2]	2.82	LYS 187[O]
2	ASN 10[ND2]	2.66	GLN 188[O]
3	GLU 407[N]	2.95	ASP 259[O]
4	ARG 483[NH1]	3.16	GLU 395[OE1]
5	ARG 483[NH2]	2.79	GLU 395[OE2]
6	LYS 387[NZ]	2.97	GLU 395[OE2]
7	SER 422[OG]	2.56	GLU 398[OE1]
8	GLY 428[N]	3.00	GLU 398[OE2]
9	LYS 387[NZ]	2.49	GLU 398[OE2]
10	LEU 261[N]	3.00	TYR 405[O]
11	TYR 415[OH]	2.76	TYR 415[O]
12	TYR 405[OH]	2.70	ARG 416[O]
13	ARG 522[NE]	2.69	GLU 480[OE2]
14	THR 517[OG1]	2.47	THR 517[OG1]
15	THR 517[OG1]	2.83	THR 517[O]
16	THR 5[O]	2.78	ARG 199[NH1]
17	LYS 187[O]	2.69	ARG 324[NH2]
18	LYS 187[O]	2.63	GLN 11[NE2]
19	GLN 188[O]	2.59	ASN 10[ND2]
20	ASP 259[O]	2.91	GLU 407[N]
21	GLU 395[OE1]	3.03	ARG 483[NH1]
22	GLU 395[OE2]	2.81	ARG 483[NH2]
23	GLU 398[OE1]	2.56	SER 422[OG]
24	GLU 398[OE2]	2.73	LYS 387[NZ]
25	GLU 398[OE2]	2.99	GLY 428[N]
26	TYR 405[O]	3.05	LEU 261[N]
27	TYR 415[O]	2.82	TYR 415[OH]
28	ARG 416[O]	2.69	TYR 405[OH]
29	GLU 480[OE2]	2.67	ARG 522[NH2]
30	GLU 484[OE1]	2.78	GLY 520[N]
31	THR 517[O]	2.87	THR 517[OG1]
32	GLY 519[O]	2.61	ARG 469[NH1]

Table II.5. List of putative hydrogen bond distances between molecule A and C of AidB. These distances are calculated in PISA Server, http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

	Monomer A residue	Distance (Å)	Monomer C residue
1	TRP3 [NE1]	2.59	ASP275 [OD2]
2	VAL8 [O]	2.95	ARG77 [NH1]
3	ASP275 [OD2]	2.70	TRP3 [NE1]

Scheme II.1. Reaction scheme of AidB proposed by Hausinger *et al.*[1] AidB repairs DNA lesion by dehydrogenase activity, forming a double bond between the methyl group and the base. The Schiff's base can be hydrolyzed by water to release the methyl group as formaldehyde.

$$
\begin{array}{c}
\text{DNA} \\
| \\
H_3C-NH(base)
\end{array}
\quad\longrightarrow\quad
\begin{array}{c}
\text{DNA} \\
| \\
H_2C=N(base)
\end{array}
\quad\longrightarrow\quad
\begin{array}{c}
\text{DNA} \\
| \\
NH_2(base)
\end{array}
$$

AidB-FAD   AidB-FADH$_2$ H$_2$O   HCHO

H$_2$O$_2$ O$_2$
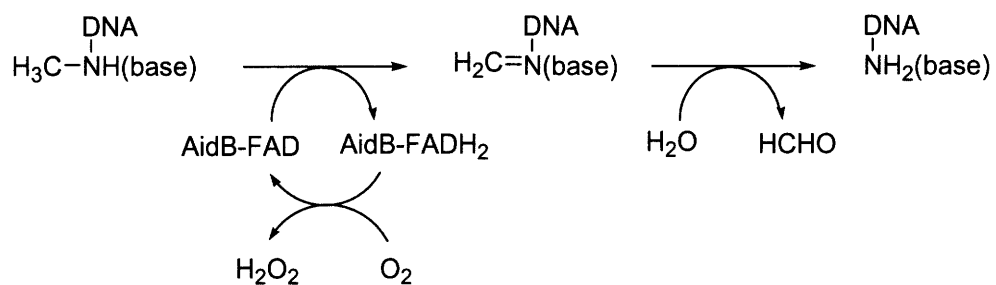
Figure II.1. A model proposed by Eichman *et al.* of AidB protecting DNA from alkylating agents by secluding the DNA helix in a 25 Å pore formed by arranging four tetrameric units around the DNA.[3] a) One tetrameric unit of AidB is colored by subunit (cyan, green, magenta, and orange), in cartoon representation. The other AidB tetramer units are shown in black cartoon. FADs in the central subunit are shown as yellow spheres. The proposed DNA binding regions of each tetramer AidB unit are colored in pink. DNA helices are shown as light grey circles. b) Side view of the same complex shown in (a) rotated 90 degree looking down the axis indicated by a black arrow. The side view shows that not all tetramer units of AidB are on the same level, and there is little interaction between each tetramer subunits that are on same level.

Figure II.1.

a
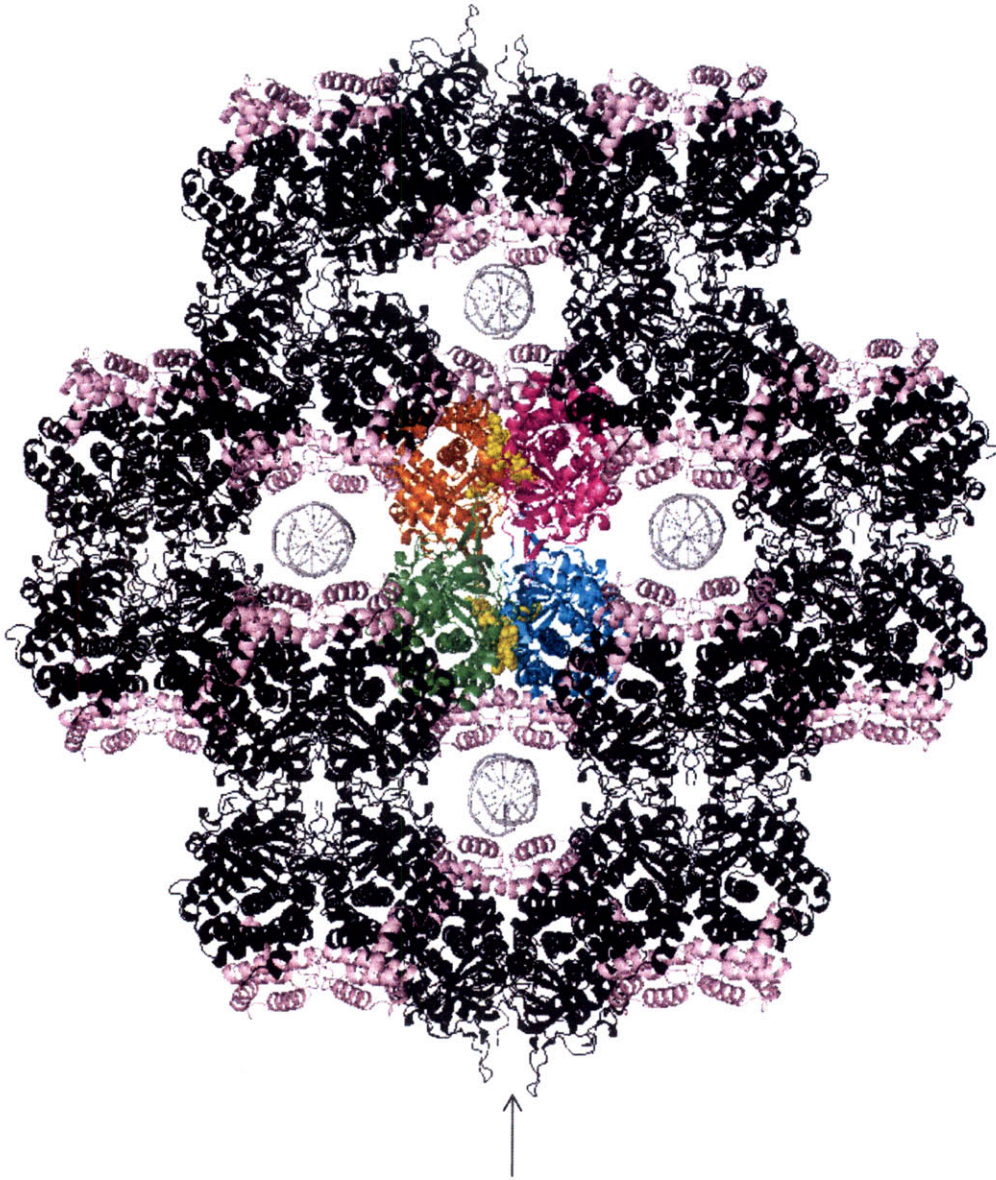
b

Figure II.2. SDS-PAGE of purified wild type AidB and triple mutant AidB. a) SDS-PAGE of purified wild type AidB. The left lane is a molecular weight marker; the molecular weight corresponding to each band is labeled on the left. The right lane is 20 µg of final purified AidB. b) SDS-PAGE of purified triple mutant of AidB. The left lane is 30 µg of final purified triple mutant AidB. The right lane is a molecular weight marker; the molecular weight corresponding to each band is labeled on the right.
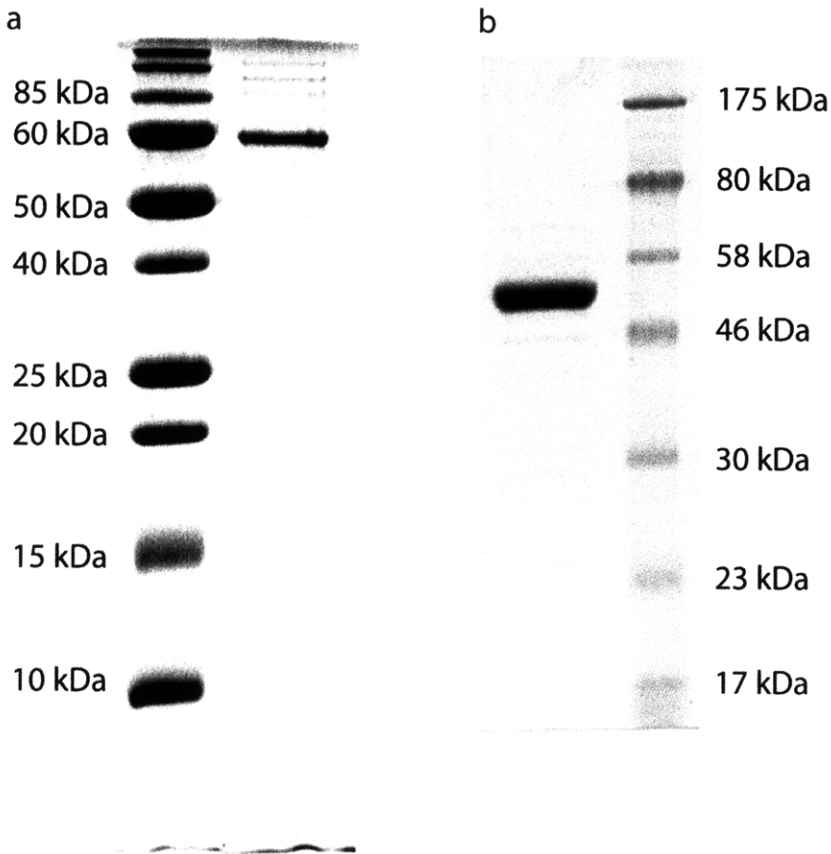
Figure II.3. Photos of native AidB crystal grown under various conditions. a) Native AidB crystals grown in 0.1 M HEPES, pH 7.5, 20% v/v ethanol, 0.25 M $MgCl_2$. The scale bar at the lower left corner is 200 μm. b) Native AidB crystals grown in 0.1 Tris, pH 8.0, 2% v/v tacsimate, 20% w/v PEG 3,000. The scale bar at the lower left corner is 200 μm. c) Native AidB crystals grown in 0.1 M imidazole, pH 8.0 0.3 M Li_2SO_4, 20% w/v PEG 3,000. The scale bar at the lower left corner is 200 μm. d) Upon streak seeding, SeMet AidB crystals grown in 0.1 M HEPES, pH 7.5, 19% v/v ethanol, 0.25 M $MgCl_2$. The scale bar at the lower left corner is 50 μm.

Figure II.3.
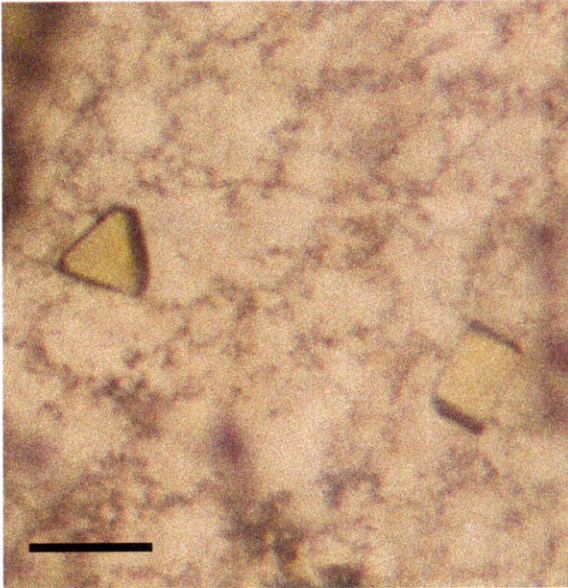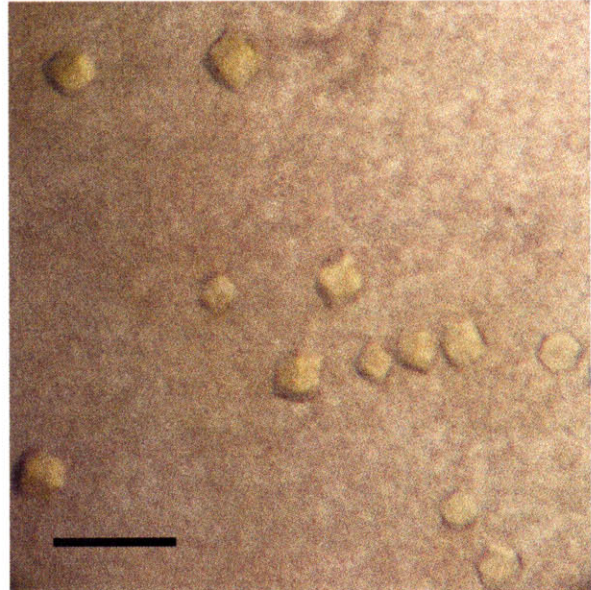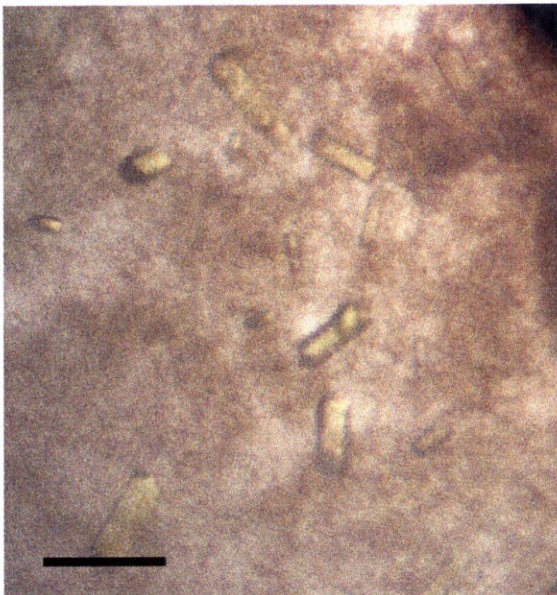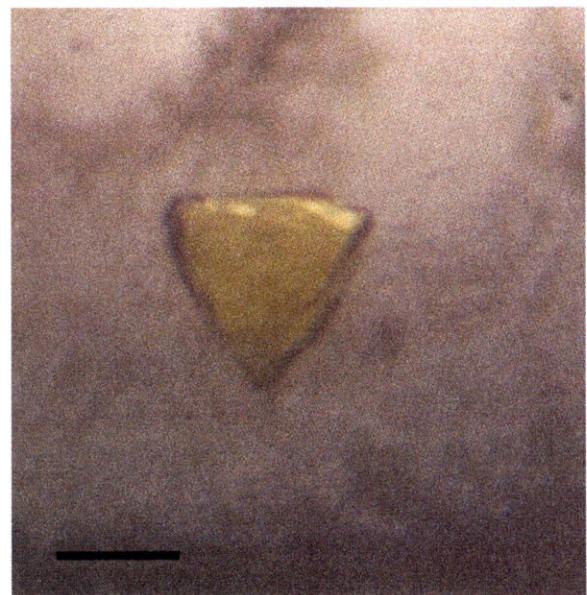
Figure II.4. Overall structure of AidB monomer shown in stereo view. The α-helical N-terminal domain (residues 25-64 and 86-161) is colored in light orange, the β-sheet domain (residues 198-282) in the middle is colored in green, a part of the α-helical C-terminal domain (residues 287-443) is colored in light cyan, another part of the α-helical C-terminal domain, the DNA binding domain proposed by Eichman *et al.*, is colored in light pink (residues 445-535), and the extra β-strands (residues 68-83) are colored in red. The N-terminus is an extended loop. The isoalloxazine ring of FAD (shown in sticks representation) binds between the α-helical region and the β-sheet region of the monomer. The adenine end of the FAD binds at the interface of two monomers (the other monomer is not shown).
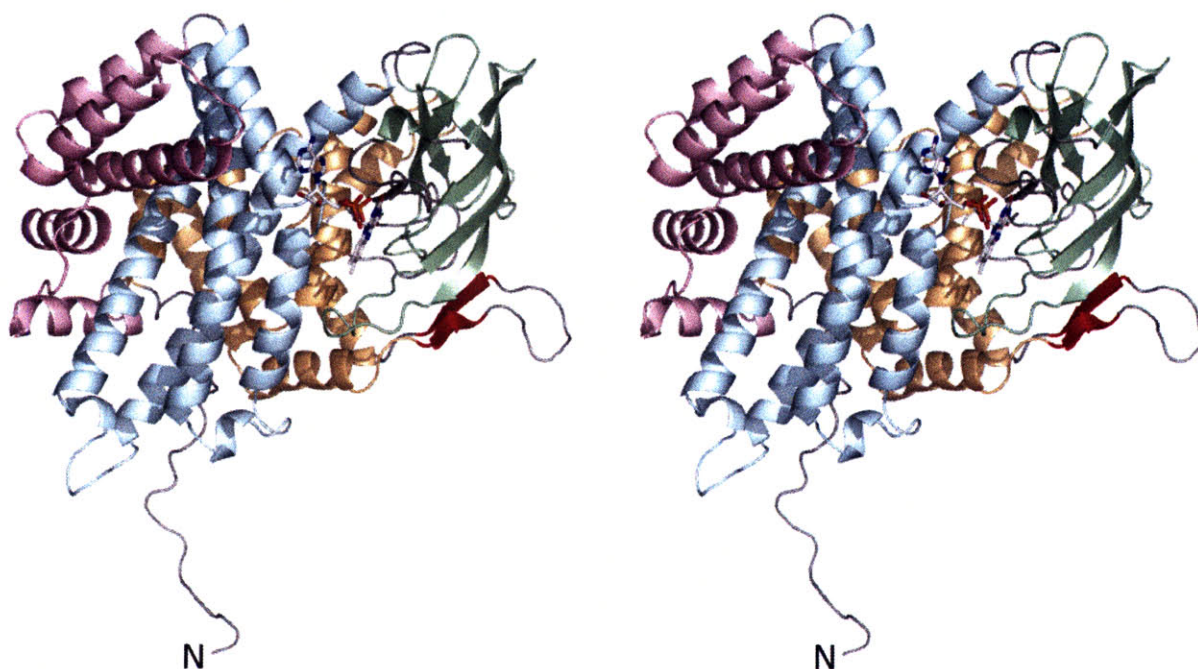
Figure II.5. Overall structure of AidB dimer colored by molecule. Molecule A is color in cyan, and molecule B is colored in magenta. a) Dimer arrangement of AidB in cartoon representation. FAD is represented in sticks. b) Dimer arrangement of AidB in surface representation. c) Dimer arrangement of AidB in ribbon representation, and FADs are shown as yellow spheres.
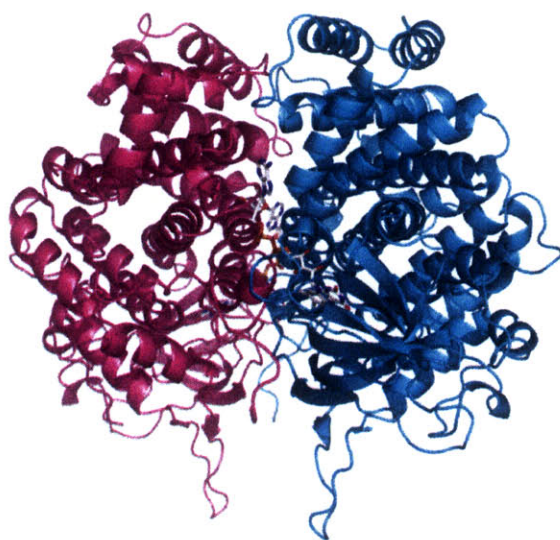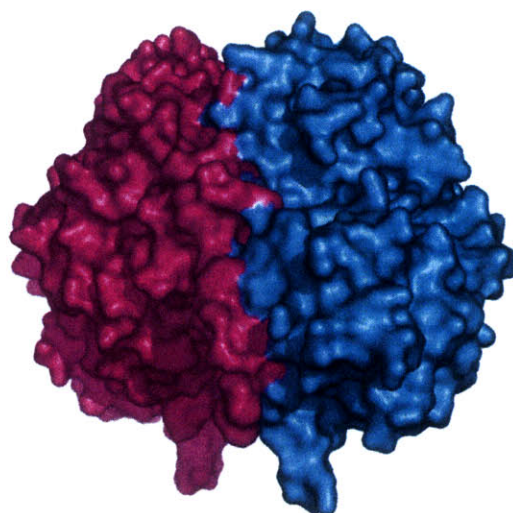
Figure II.6. Structural alignment of AidB, isovaleryl CoA dehydrogenase (IVD, pdb code 1IVH), and medium chain acyl CoA dehydrogenase (MCAD, pdb code 3MDE). Secondary structure of AidB is defined above the AidB sequence, and the secondary structure of IVD is defined below its sequence. Residues that are strictly conserved across all three proteins are highlighted in red. Alpha helices are indicated by helical lines with α labels, beta strands are indicated by arrows with β labels, and 3_{10} helices are indicated by helical lines with η labels.

Figure II.6.


```
                                                  α1          η1        α2                    α3
                                               ℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓ          ℓℓℓℓℓℓℓℓℓ
              1         10        20        30          40        50          60
AidB  ......VAWQTHTVPNQPIPLNNSNLYLSDGALCEAVTREGAGWDSLPLASIGQQLGTAESLELGRLAN
MCAD  .........................................GPSPELTEEQQEFQATAREFARE.EIIPVAAEYD
IVD   MATATRLLGWRVASWRLRPPLAGFVSQRAHSLLPVDDAINGLSEEQRQLRQTMAKFLQE.HLAPKAQEID
                                                 ℓℓℓℓℓℓℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ ℓ ℓ   ℓℓℓℓℓ
                                                    η1          α1                     α2

           β1              β2          α4             η2                          α5
       ▶▶▶▶▶▶▶        ▶▶▶▶▶▶▶     ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ    ℓℓℓ              ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
  ℓℓ      70        80        90        100       110         120
AidB  VNFPELLRYDAQGERLDEVRFHPAWHLLEQALCTNEVHNLAWEE.....DARSGAFVAKAARFMLHAQVE
MCAD  RTGE...............YPVPLKRAWELGLMNTHIPESFGGLGL...GIIDSCLITEELAYGC
IVD   RSNE...............PKNLREFWKQLGNLGVLGITAPVQYGGSGL...GYLEEVLVMEELSRAS
  ℓℓ                   ℓℓℓℓℓℓℓℓ         ℓℓℓ       ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
                          α3              η2                    α4

            α6              η2    α7           η4       β3          η5
       ℓℓℓℓℓℓℓℓℓℓℓ  ℓℓℓℓℓℓℓℓℓ ℓℓℓ ℓℓℓℓℓℓℓℓ       ℓℓℓ    ▶▶▶▶▶       ℓℓℓ
     130       140       150       160       170       180       190
AidB  AGSLCPITMTF.AATPLLL.QMLPAPFQLWTTELLSDRYDSHLLPGGQKRGLLIGMGMTEKQGGSDVMSN
MCAD  TGVQTAIEANT.LGQVPLIIGGNYQQQKKYLGRMTE.............EPLMCAYCVTEPGAGSDVAGI
IVD   GAVGLSYGAHSNLCINQLVRNGNEAQKENYLPKLIS.............GEYIGALAMSEPNAGSDVVSM
      ℓℓℓℓℓℓℓℓℓℓ   ℓℓℓℓℓℓℓℓℓ ℓℓℓℓℓℓℓℓℓ           ▶▶▶▶▶▶▶        ℓℓℓ
          α5           α6           α7                  β1                 η3

        β4              β5              β6             β7                 β8
     ▶▶▶▶▶         ▶▶▶▶▶▶▶         ▶▶▶▶▶         ▶▶▶▶▶▶▶▶           ▶▶▶▶▶▶
    200       210       220       230       240       250
AidB  TTRAERLEDGSYRLVGHKWPFSVP.QSDAHLVLAQTAG.......GLSCFVPRPLPDGQRNAIRLERL
MCAD  KTKAEKKGD.EYIINGQKMWITNGGKANWYFLLARSDPDPKAPASKAFTGFIVEAD....TEGVQIGRK
IVD   KLKAEKKGN.HYILNCNKPWITNGPDADVLIVYAKTDLAAVPAS.RGITAFIVEKG....MEGFSTSKE
      ▶▶▶▶▶         ▶▶▶▶▶▶▶         ▶▶▶▶▶         ▶▶▶▶▶▶▶▶           ▶▶▶▶▶▶
        β2              β3              β4             β5                 β6

                   β9             α8    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        α9
       ▶▶▶▶▶▶▶▶▶▶▶ · ·▶▶▶▶    ℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ       ℓℓℓℓℓℓℓℓℓℓℓℓ
    260       270       280       290       300       310       320
AidB  KDKLGMRSNASCEVEFQDAIG...WLLGLEGEGIRLILKNGGMTRPDCALGSHAMMRRAPSLAIYHAHQR
MCAD  EINMGQRCSDTRGIVFEDVRVPKENVLTGEGAGFKIAMGTFDKTRPPVAAGAVGLAQRALDEATKYALER
IVD   LDKLGMRGSNTCELIFEDCKIPAANILGHENKGVYVLMSGLDLERLVLAGCPLGLMQAVLDHTIPYLHVR
      ▶▶▶▶▶▶▶▶▶▶▶▶▶    ℓℓℓ ℓℓ     ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓ
          β7         η6 β8               α8                  α9

    β10  311 η6           α10                              α11
    ▶▶▶  ▶▶▶ ℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
        330       340       350       360       370       380       390
AidB  HVFGNPLIQQPLMRHVLSRMALQLEGQTALLFRLARAWDPRADAKEALWARLFTPAAKPVICKRGMFFVA
MCAD  KTFGKLLAEHQGISFLLADMAMKVELARLSYQRAAWEIDSGR.....RNTYYASIAKAYAADIANQLAT
IVD   EAFGQKIGHFQLMQGKMADLYTRLNACRQYVYNVAKACDEGH.....CTAKDCAGVILYSAECATQVAL
      ▶▶▶  ▶▶▶ ℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
       β9  310 η7              α10                      α11

          η7          α12        η8     α13           α14
     ℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓ ℓℓℓℓ ℓℓℓℓℓℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓℓ      ℓℓℓ
      400       410       420       430       440       450       460
AidB  EAMEVLGGIGYCEESELERLYREMPVNSIWEGSGNEMCLDVLRVLNKQASVYDLLSEAPVEVKGQDRYFD
MCAD  DAVQVFGGNGFNTEYPVEKLMRDAKIYQIYECTAQIQRIIIAREHIGRYK...............
IVD   DGIQCFGGNGYINDFPMGRFLRDAKLYEIGACTSEVRRLVIGRAFNADFH...............
     ℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓ       ℓℓℓℓℓℓℓℓℓℓℓℓ
        η8            α12                  α13

      α15          η9           α16              α17                 α18
    ℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        ℓℓℓℓℓℓℓℓℓℓℓ
      470       480       490       500       510       520       530
AidB  RAVRRLQQQLRKPAEELGREITHQLFLLGCGAQCMLKYASPPMAQAWCQVMLDTRGGVRLSEQIQNDLLLR
MCAD  ..................................................................
IVD   ..................................................................

           540
AidB  ATGGVCV
MCAD  ......
IVD   ......
```
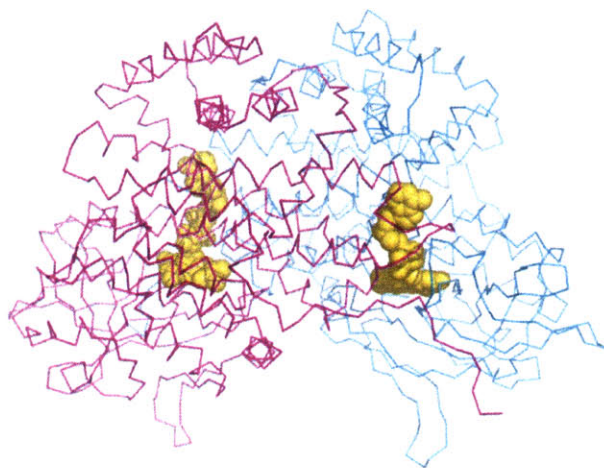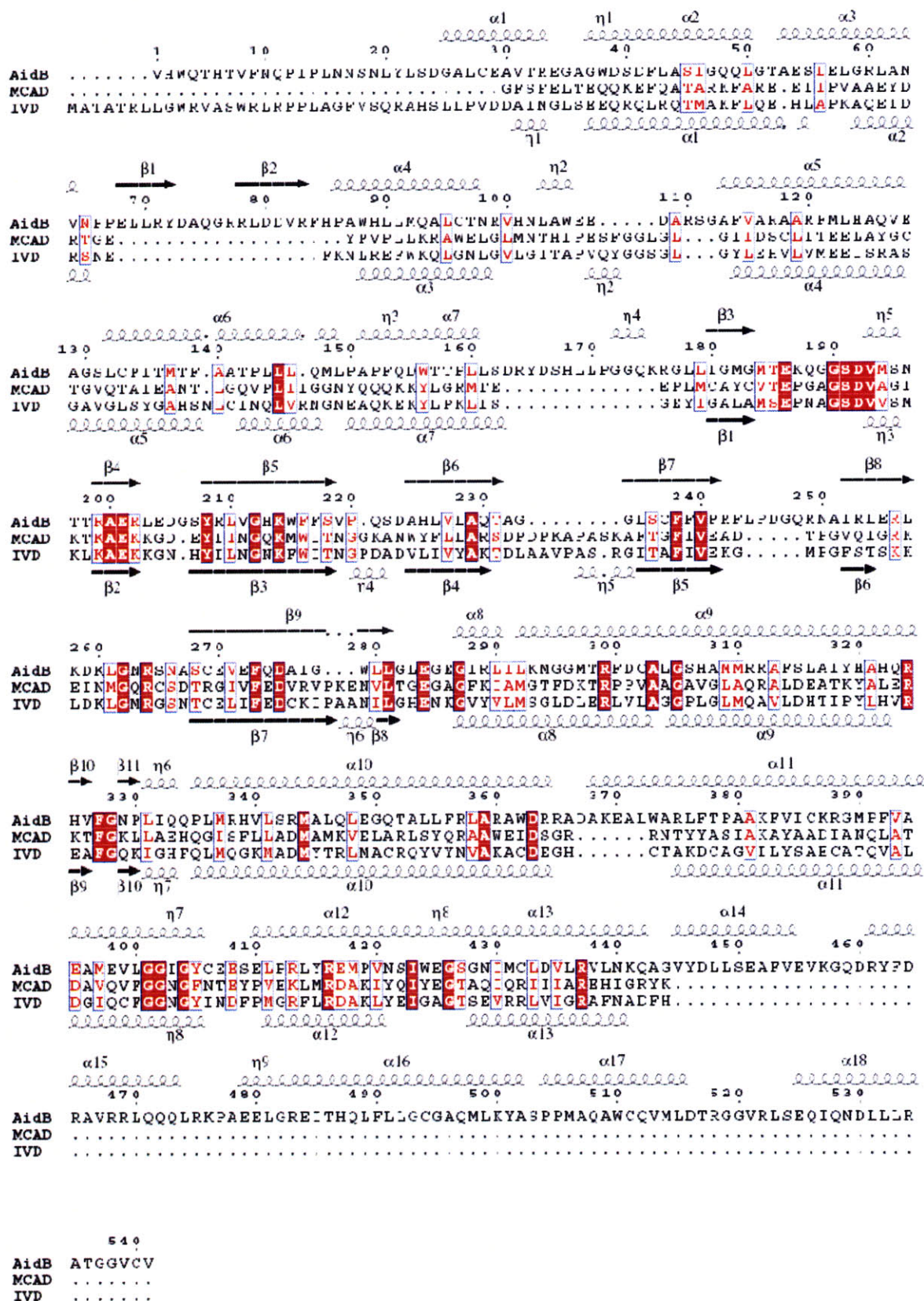
Figure II.7. Structural overlay of AidB, isovaleryl CoA dehydrogenase (IVD, pdb code 1IVH), and medium chain acyl CoA dehydrogenase (MCAD, pdb code 3MDE). a) Overlay of tetramer unit of AidB, with dimer unit of IVD and MCAD. Tetramer unit of AidB found in asymmetric unit is colored in cyan, dimer unit of IVD is colored in pink, and dimer unit of MCAD is colored in orange. The top bracket denotes the extra α-helices in AidB that are not present in IVD or MCAD. The middle bracket denotes the AB dimer unit of AidB, IVD and MCAD. The bottom bracket denotes the CD dimer of AidB. b) Overlay of tetramer unit of AidB and tetramer unit of IVD. AidB is colored in cyan, and IVD is colored in pink. The top bracket denotes the CD dimer of IVD. The middle bracket denotes the AB dimer of AidB and IVD. The bottom bracket denotes the CD dimer of AidB.
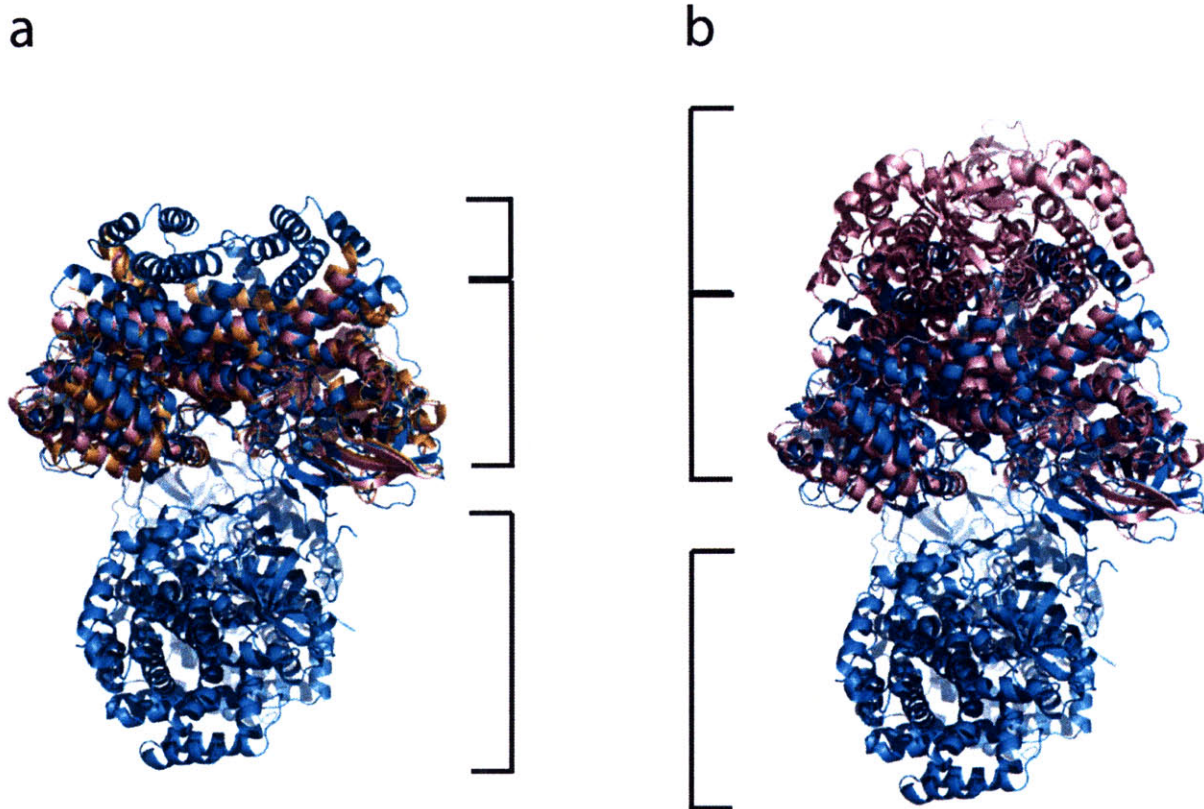
Figure II.8. Electrostatic surface of the putative substrate binding tunnel in AidB. a) Cut-away view of the putative substrate binding tunnel in AidB. The tunnel surface is shown in electrostatic representation; red indicates negatively charged residues on the surface, and blue indicates positively charged residues on the surface. FAD is shown in yellow stick representation; N3 nitrogen of the isoalloxazine ring is labeled. W424 is shown in surface representation, and colored light green for clarity. The red patch at the end of the tunnel is where Glu425 located. The tunnel is 25 Å in length, and 6 – 15 Å in width. b) Top view of the putative substrate binding tunnel. The protein surface is shown in electrostatic representation; red indicates negatively charged residues on the surface, and blue indicates positively charged residues on the surface. FAD is shown in stick representation. There is a cleft with positively charged residues on the surface around the entrance to the tunnel. The scale bar at the bottom is about 10 Å. c) Superposition of the putative substrate binding tunnel in AidB, and substrate binding site in IVD and MCAD. The tunnel surface of AidB is shown in electrostatic representation; red indicates negatively charged residues on the surface, and blue indicates positively charged residues on the surface. FAD of AidB is shown as yellow sticks, FAD of IVD is shown as green sticks, and FAD of MCAD is shown as red sticks. CoA persulfide in the substrate binding site of IVD is shown as cyan sticks, and octanoyl CoA in the substrate binding site of MCAD is shown as magenta sticks.

Figure II.8.

a
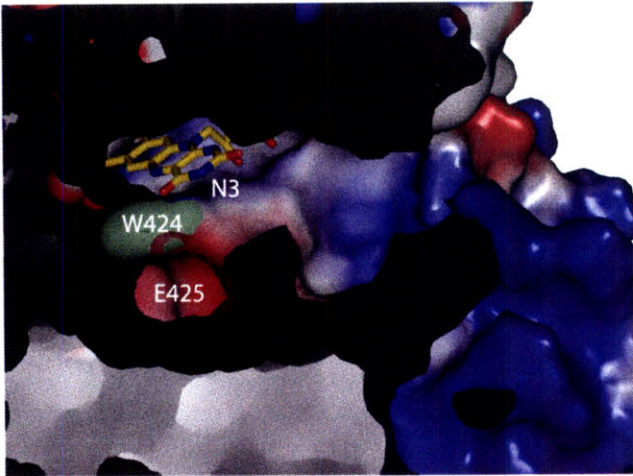
b
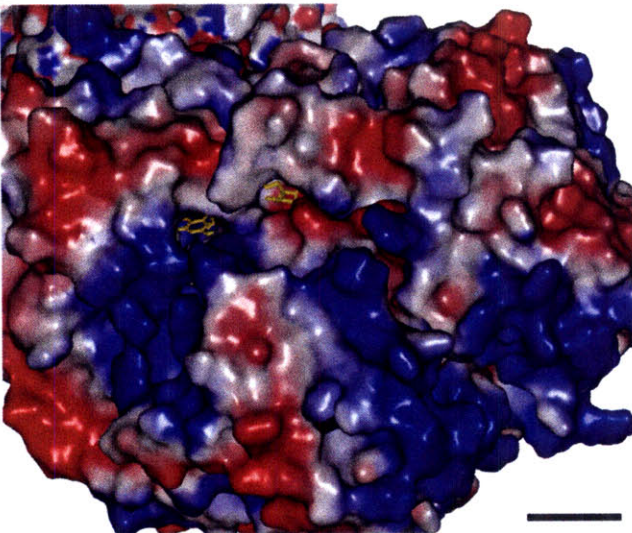
c

Figure II.9. Residues near the FAD binding site in AidB. a) Distances between T185, S191, S218, R324 and FAD are shown. b) Distances between C133, 302, 386, and between N5 of FAD and C133 are shown.
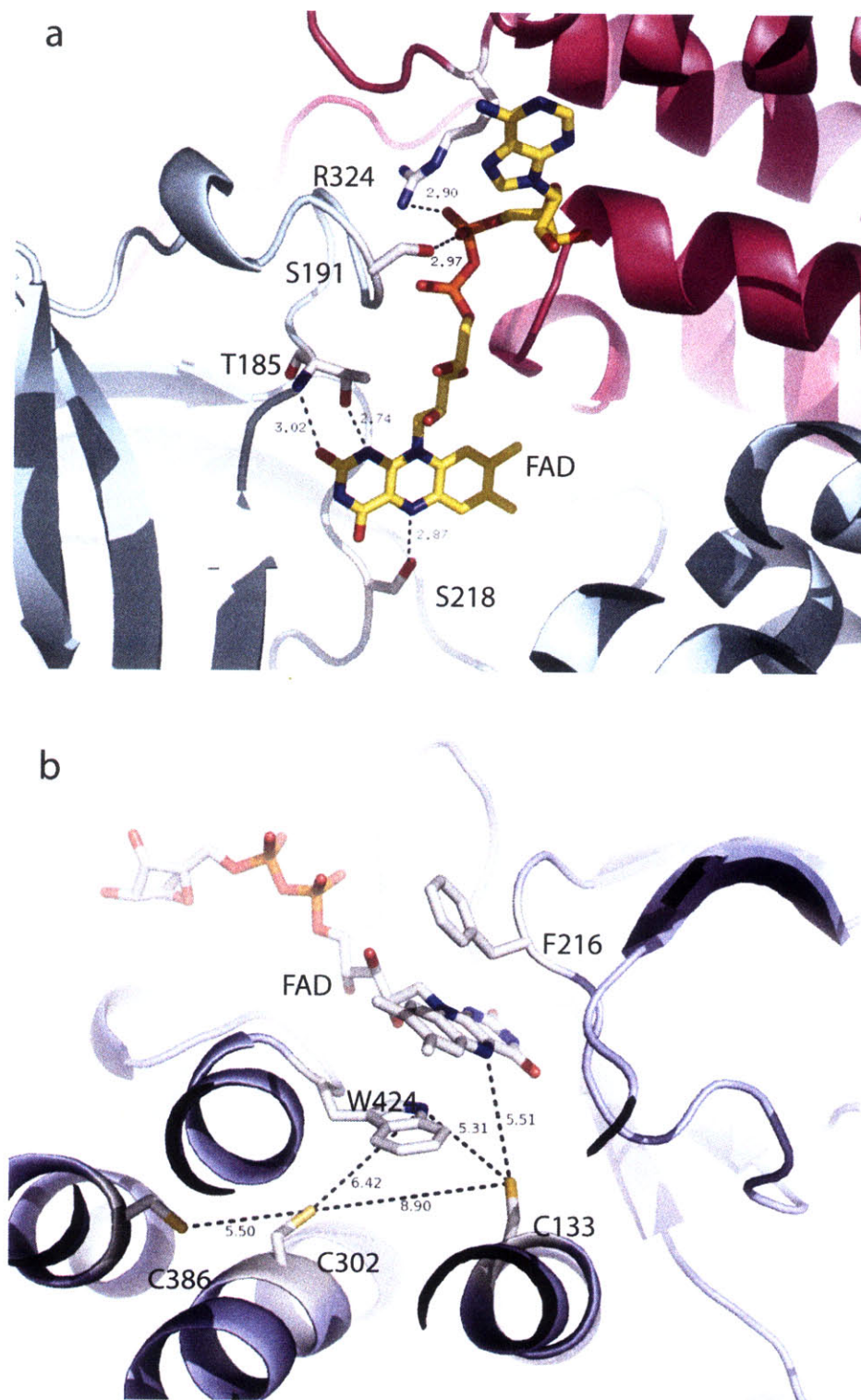
Figure II.10. Electrostatic surface representation of the AB dimer of AidB. Red indicates negatively charged residues on the surface, and blue indicates positively charged residues on the surface. The putative substrate binding tunnel is shown in green spheres. a) DNA binding domain proposed by Eichman *et al.* shown in electrostatic surface representation.[3] Yellow stars indicate the proposed DNA binding residues. Black arrows indicate the entrance to the putative substrate binding tunnel. b) Side view of the AB dimer that shows the DNA binding domain proposed by Eichman *et al.* is 38 Å from the entrance to the putative substrate binding tunnel. Yellow stars indicate the proposed DNA binding residues by Eichman.[3] Black arrow indicates the shortest trajectory on the surface of the protein that goes from the middle of the positively charged patch of the proposed DNA binding domain to the entrance of the tunnel

a

b

R475

R468

R518 · R518

R468

R475

38 Å
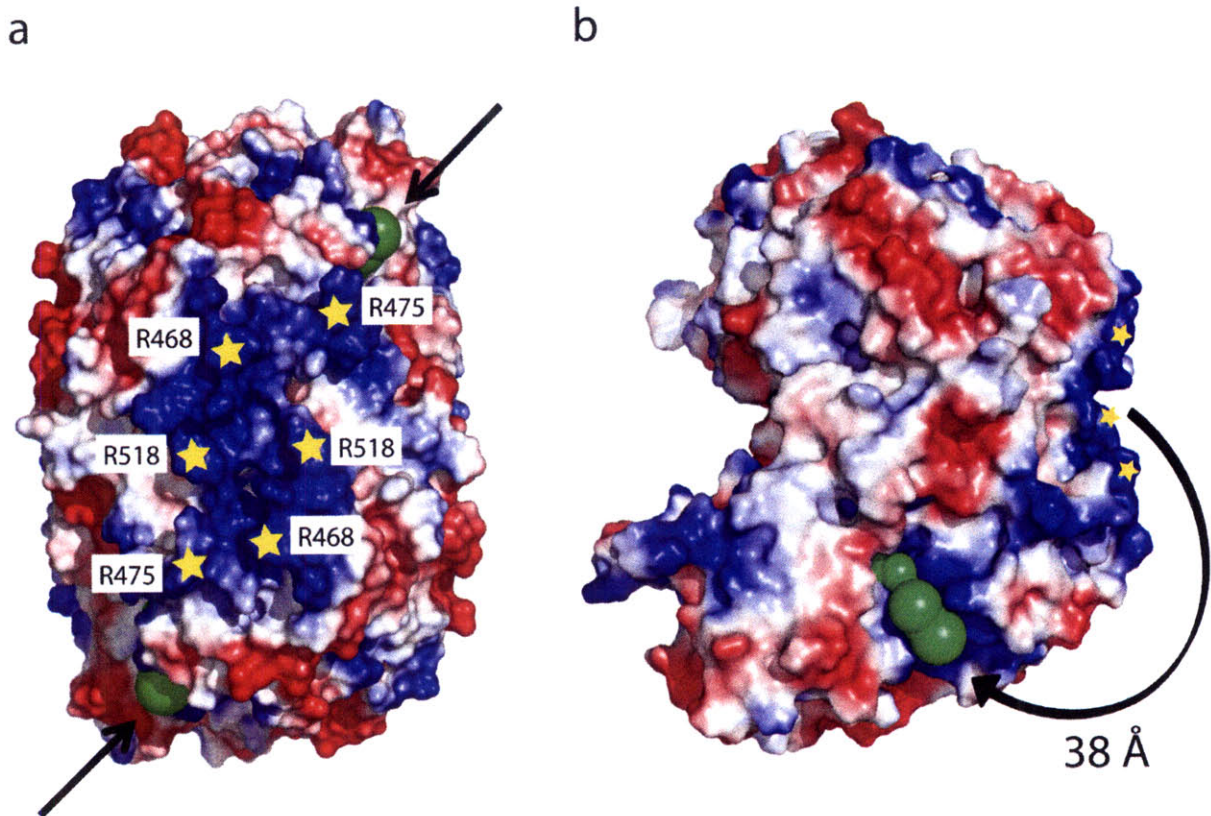
Figure II.11. Sliced view of the docking model of an ethenoadenine flipped out of a dsDNA inside the putative substrate binding channel in AidB. Double stranded DNA is shown in light green cartoon representation. The flipped out ethenoadenine base and FAD are shown in stick representation. The shortest distance between the ethenoadenine and the N5 nitrogen of the isoalloxazine ring is labeled.
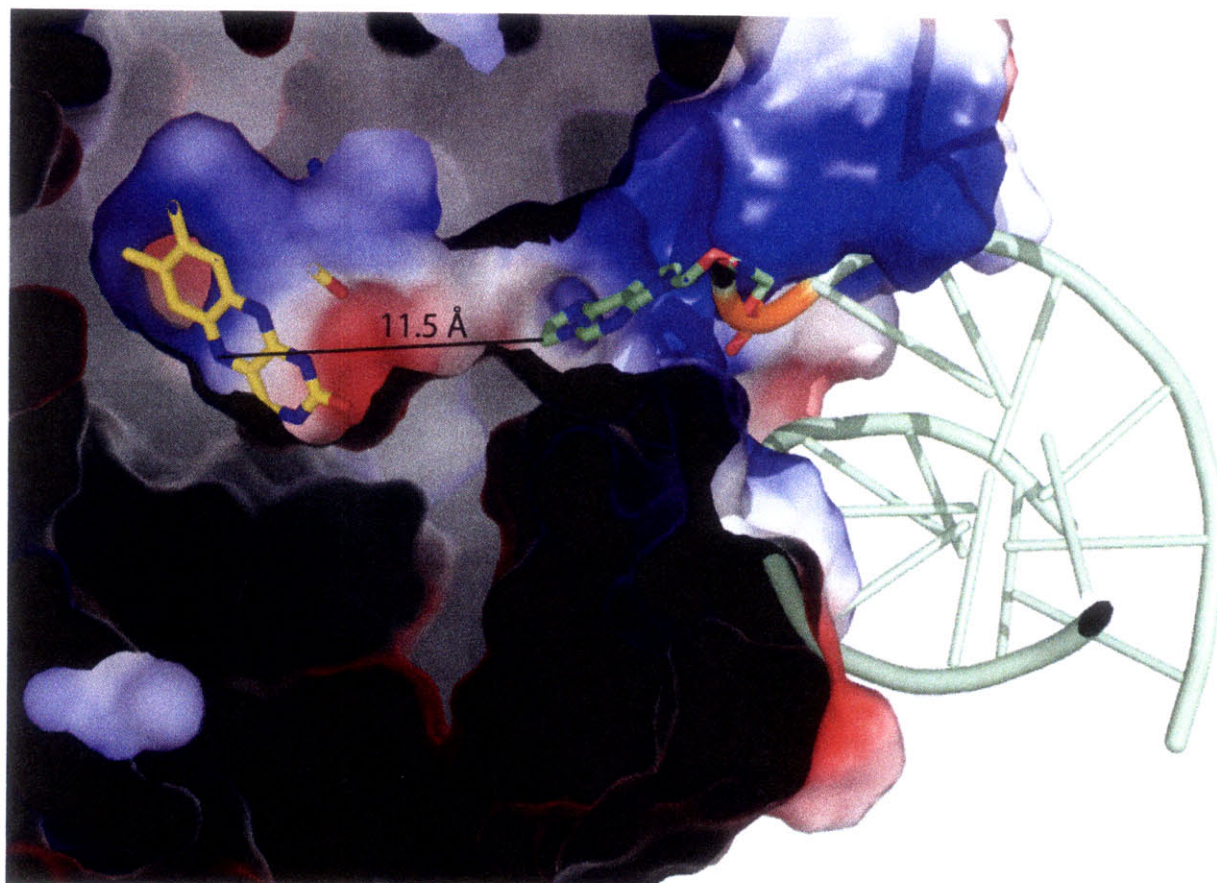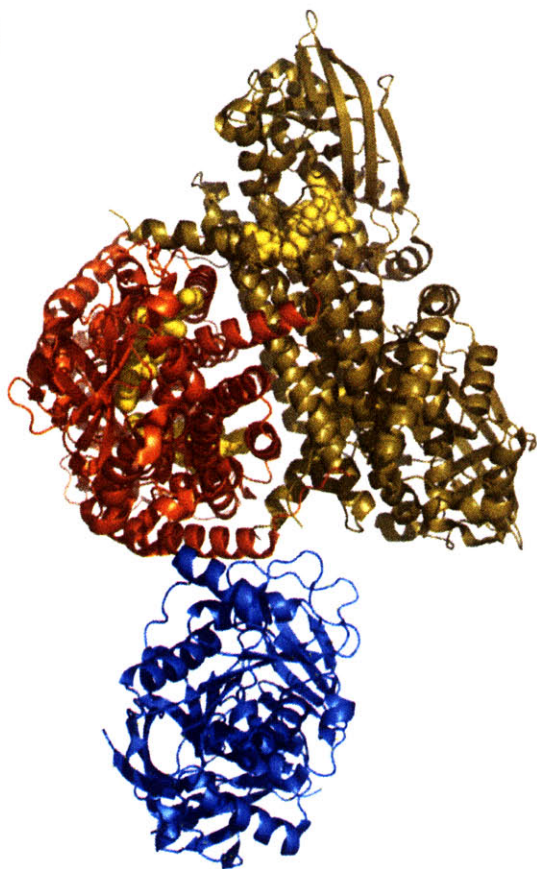
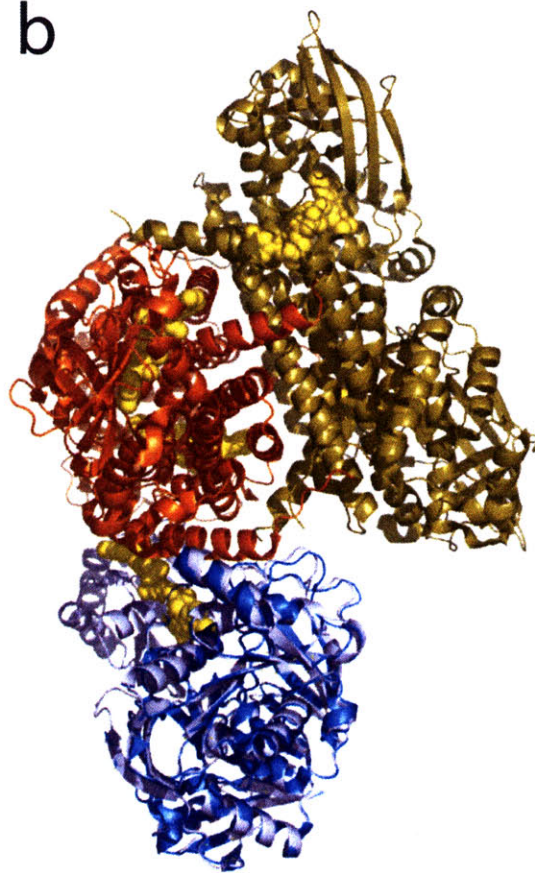**Figure II.12.** Docking model of AidB tetramer with human ETF based on MCAD-ETF structure. a) Crystal structure of MCAD-ETF (1T9G) with ETF FAD domain disordered.[21] One MCAD dimer is colored in orange, another MCAD dimer is colored in deep olive. ETF is colored in blue. FAD bound to MCAD is shown as yellow spheres. b) Superposition of full human ETF (1EVF) structure solved separately with the structure of MCAD-ETF complex. One MCAD dimer is colored in orange, another MCAD dimer is colored in deep olive. The ETF from the crystal structure of the MCAD-ETF complex is colored in blue. The full ETF structure determined separately is colored in light blue. FADs are shown in yellow spheres. c) Docking model of AidB tetramer with the human electron transfer flavoprotein (ETF) based on the superposition of molecules A and B of AidB with orange dimer of MCAD. With this superposition, ETF clashes with the second dimer of AidB (molecules C and D). Molecule A is colored in cyan, molecule B is colored in magenta, molecule C is colored in yellow, and molecule D is colored in green. The human ETF is colored in light blue, shown in surface representation.
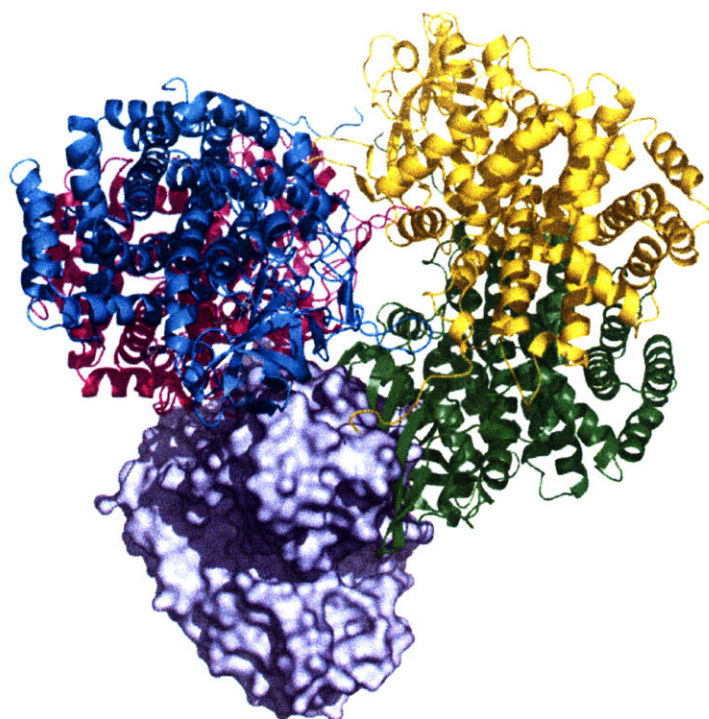
Figure II.12.

Figure II.13. Docking model of AidB dimer with human ETF. a) Molecule A of AidB is colored in cyan, molecule B of AidB is colored in magenta, and the ETF is colored in light blue. FAD's from both AidB and ETF are colored in yellow, shown in stick representation. Minor clash between the FAD binding domain of ETF and the N-terminus of monomer B of AidB is circled. b) The AidB dimer is shown in electrostatic surface representation. The human ETF is colored in light blue, and the "recognition loop" (β91-β200) is colored in black. c) Molecule A of the AidB dimer is colored in cyan, molecule B of the AidB dimer is colored in magenta, and the human ETF is colored in light blue. FAD's from molecule A of AidB and from ETF are colored in yellow, and the distance between them is labeled.
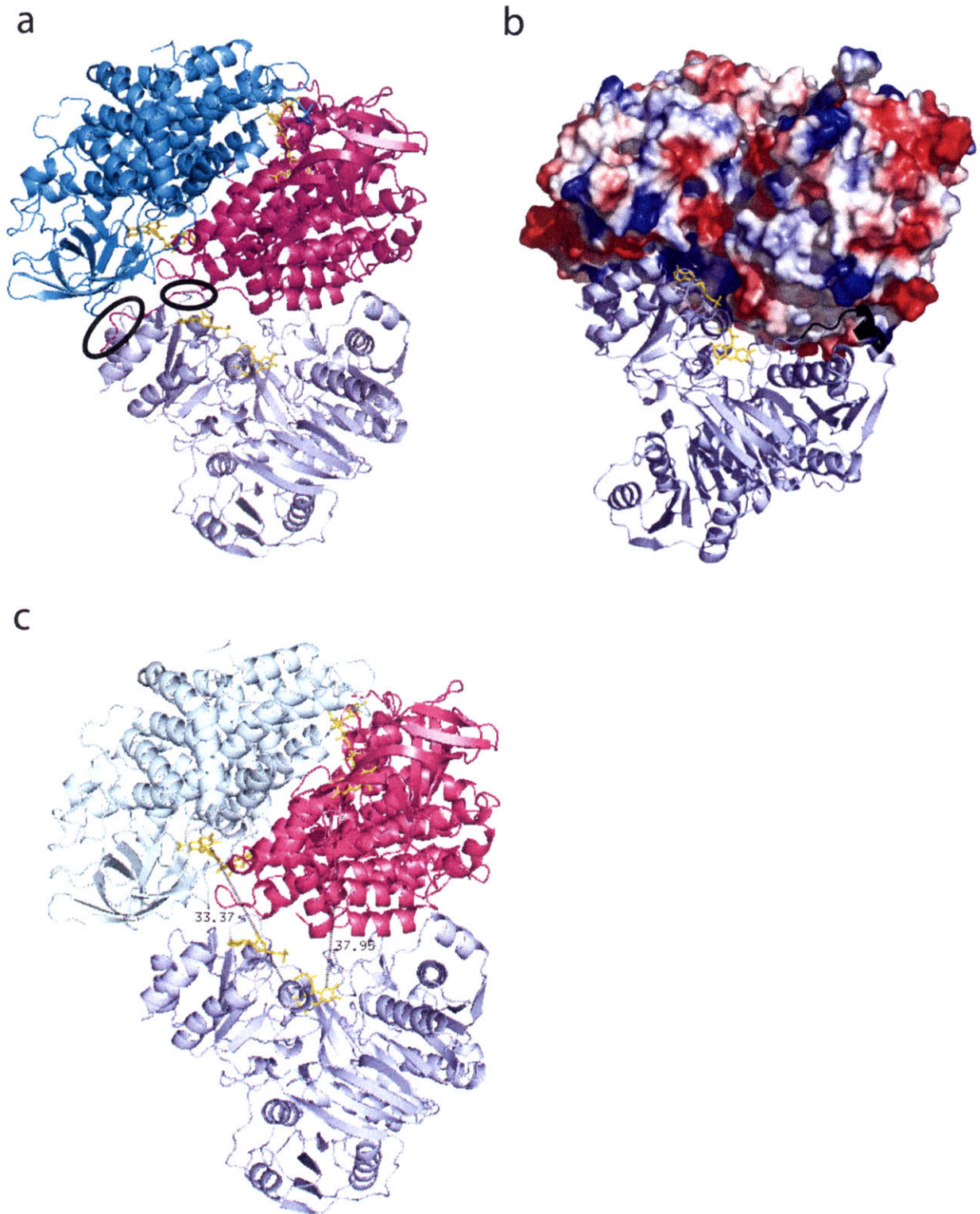
Figure II.13.

a

b

c

33.37

37.90

114

Figure II.14. Sedimentation velocity of 6.3 μM T185V/S191R/R324D triple mutant of AidB. The g(s*) plot is shown on the top, and the residual plot is shown at the bottom. The experimental g(s*) data points are shown as blue squares (■) and the fitted g(s*) plot is shown as a continuous red line. The residual is best fit to 9.64E-3. Sedimentation velocity of 17.8 μM triple mutant AidB yielded similar result (data not shown).
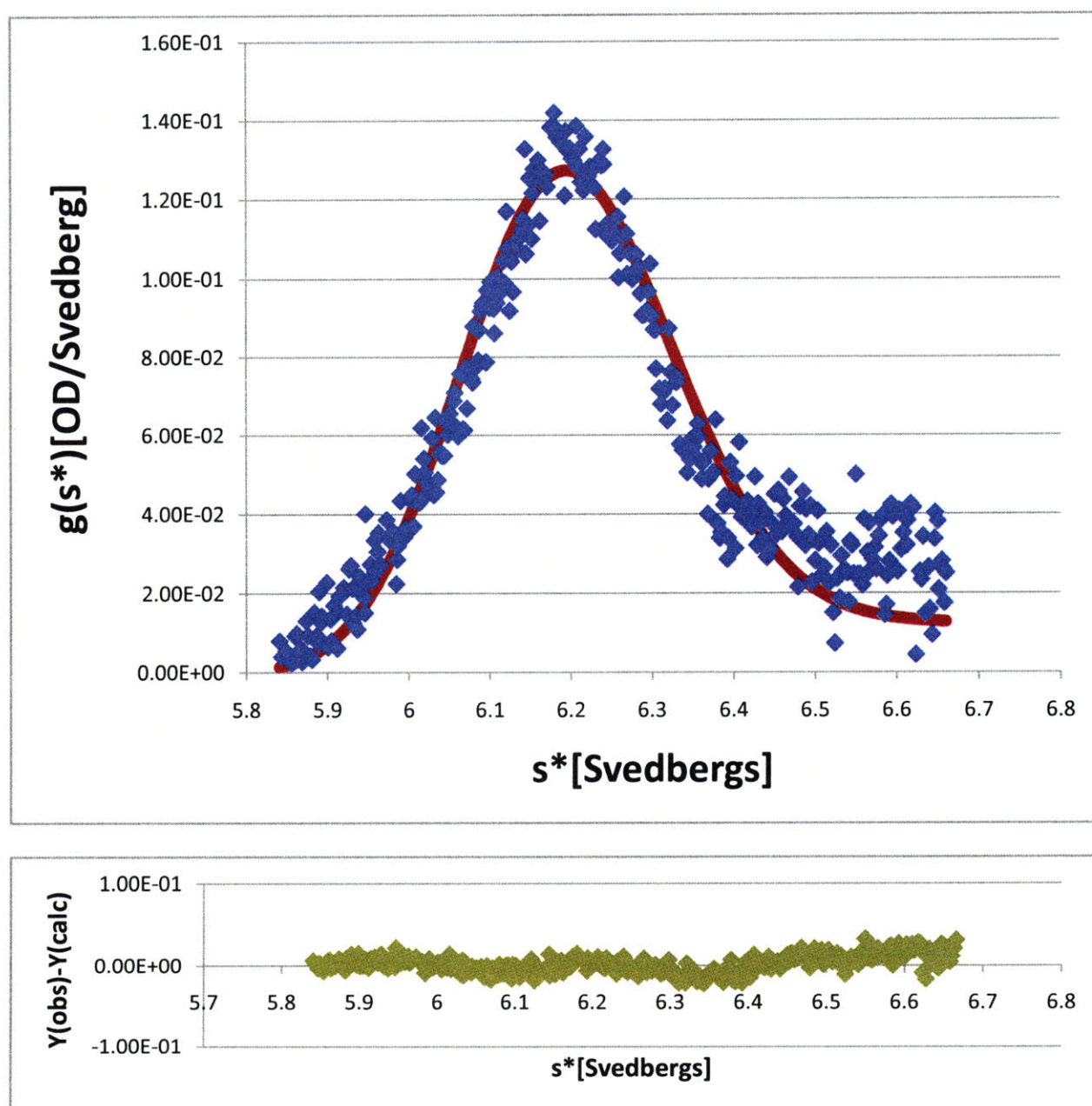
Figure II.15. Sedimentation velocity of 3 µM apo (FAD-free) AidB. The g(s*) plot is shown on the top, and the residual plot is shown at the bottom. The experimental g(s*) data points are shown as open red circles (○) and the fitted g(s*) plot is shown as a continuous blue line. The residual is best fit to 2.40E-2.
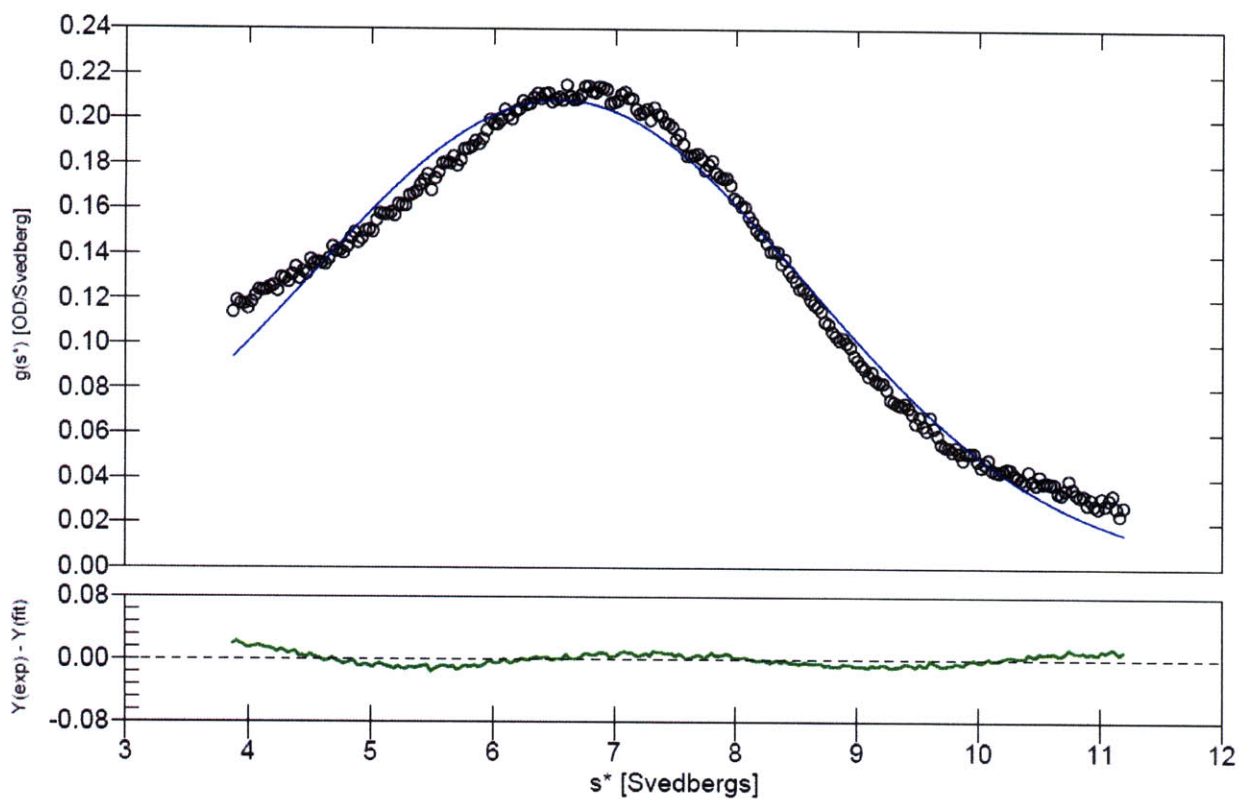
Figure II.16. Overall crystal packing of AidB in the asymmetric unit in P3$_2$ space group. The AidB molecules are colored by tetrameric units. FADs are shown in spheres representation.
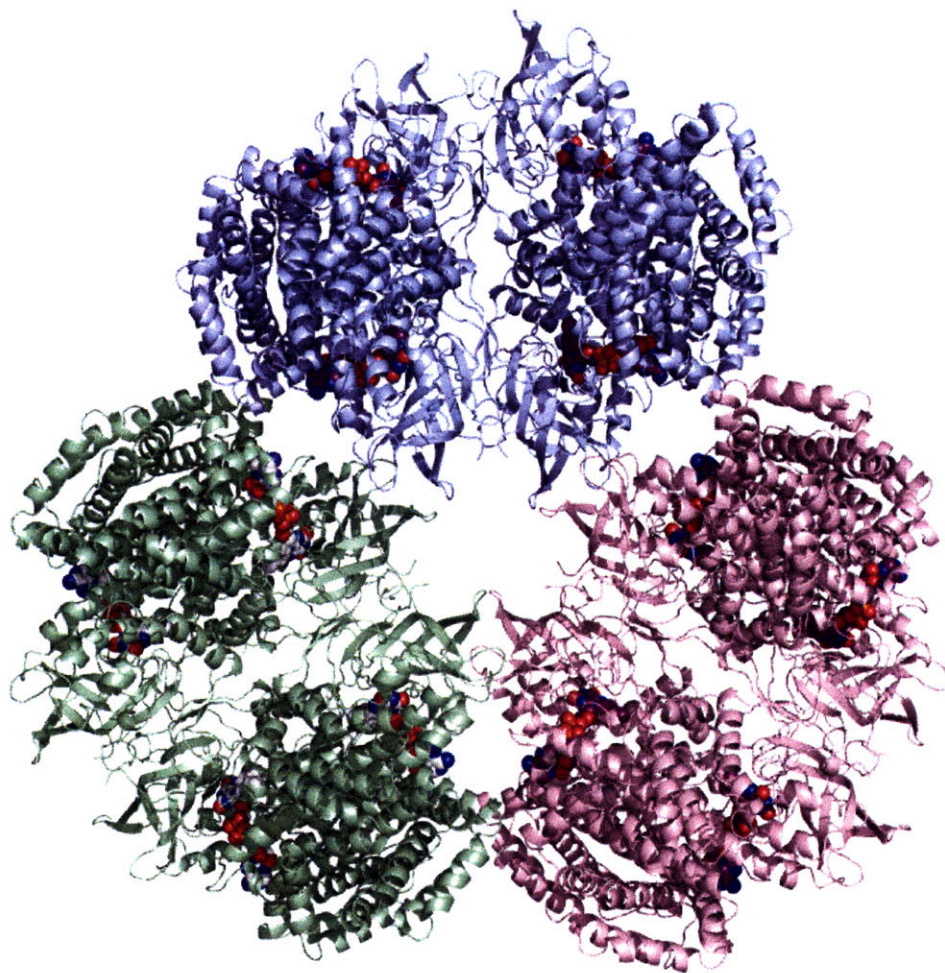
Figure II.17. Overview of the tetramer unit of AidB. Molecule A is color in cyan, molecule B is colored in magenta, molecule C is colored in yellow, and molecule D is colored in green. a) Tetramer arrangement of AidB in cartoon representation. FAD is represented in sticks. b) Tetramer arrangement of AidB in surface representation.

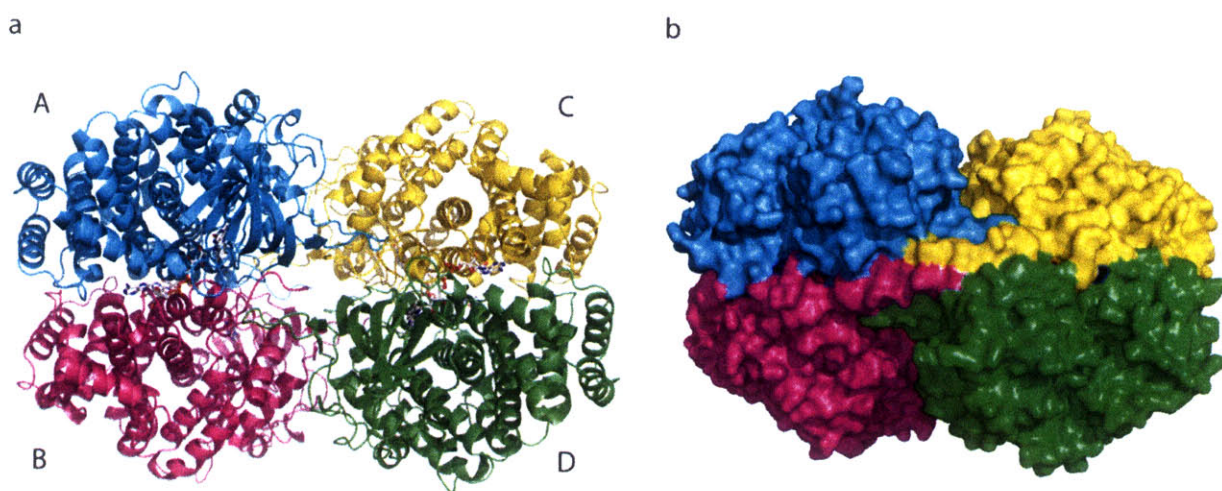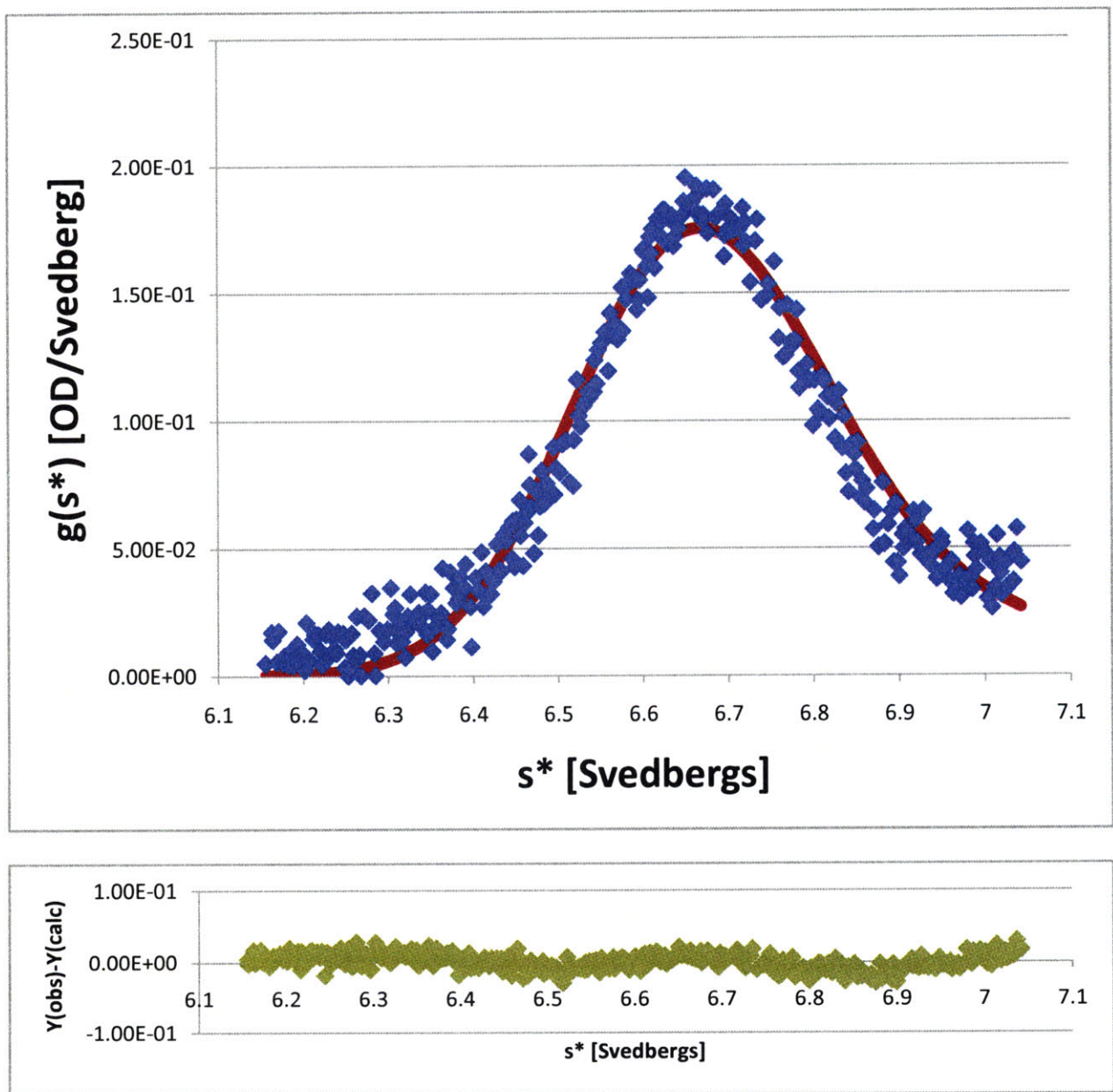**Figure II.18.** Sedimentation velocity of 6.1 μM wild type AidB. The g(s*) plot is shown on the top, and the residual plot is shown at the bottom. The experimental g(s*) data points are shown as blue squares (■) and the fitted g(s*) plot is shown as a continuous red line. The residual is best fit to 9.38E-3. Sedimentation velocity of 7.5 μM wild type AidB yielded similar result (data not shown).

Chapter III

Characterization of AlkB and hABH 2-3

III.A. Summary

E. coli AlkB is a DNA repair enzyme that belongs to the α-ketoglutarate (αKG) mononuclear Fe(II)-dependent dioxygenase family. AlkB directly repairs DNA damage via oxidative demethylation. There are eight different AlkB homologues in humans (hABH1-8); however, the activities of only three of the homologues (hABH1-3) are known.[1, 2] The substrate preference for AlkB and its functionally known human homologues differ greatly. Specifically, AlkB repairs N1-methyladenine (1mA) and N3-methylcytosine (3mC) lesions on both DNA and RNA, whereas hABH2 repairs 1mA and 3mC on DNA only, and hAHB3 repairs both DNA and RNA.[1, 2] The crystal structure of AlkB with methylated DNA substrate (TmAT) was recently solved.[3] This chapter describes, a new purification scheme developed to purify large quantity of AlkB protein to > 95% purity and free of exonuclease activity, which is important for crystallization and biochemical studies. Our collaborators in Professor John Essigmann's group used the AlkB protein purified with this new purification scheme and found that AlkB also repairs alkylated lesions: $1,N^6$-ethenoadenine (ϵA) and $1,N^6$-ethanoadenine (EA) lesions.[4, 5] The purifications of hABH2 and hABH3 were also established, and have been shown to be active by Jim Delaney in Professor John Essigmann's group. In addition, the structure of the substrate-free form of AlkBΔN11 was solved in the P1 space group.

This chapter is reproduced in part from:

1. J. C. Delaney, L. Smeester, C. Wong, L. E. Frick, K. Taghizadeh, J. S. Wishnok, C. L. Drennan, L. D. Samson, J. M. Essigmann, AlkB Reverses Etheno DNA Lesions Caused By Lipid Oxidation *In Vitro* and *In Vivo*; *Nat. Struct. Mol. Biol.* (2005), 12(10), 855-60.

2. L. E. Frick, J. C. Delaney, C. Wong, C. L. Drennan, J. M. Essigmann, Alleviation of 1,N^6-Ethanoadenine Genotoxicity by the *Escherichia coli* Adaptive Response Protein AlkB; *Proc. Natl. Acad. Sci. USA* (2007), 104(3), 755-60.

Protein purifications, crystallization, and structure determination were done by C. Wong. All biochemical studies and mass spectrometry were done by Jim Delany and Lauren Frick in Professor John Essigmann's laboratory, with the exception of the MMS plate mentioned in this chapter which was done by C. Wong. Structural analysis was done by both C. Wong and C.L. Drennan.

III.B. Introduction

Both endogenous and environmental alkylating agents are ubiquitous in nature. These alkylating agents damage DNA, causing cytotoxicity and mutagenicity in cells. Organisms have various DNA repair pathways to protect themselves against alkylating agents, one of which is by direct reversal of alkylation damage. AlkB is one of the DNA repair proteins that is part of the adaptive response in *E. coli*, and it is up-regulated in the presence of methylating agents.[6-10]

AlkB belongs to the α-ketoglutarate (αKG) mononuclear iron-dependent dioxygenase family.[11, 12] Studies have shown that AlkB repairs various DNA lesions by oxidative demethylation, including N1-methyladenine (1mA), N3-methylcytosine (3mC), N1-methylguanine (1mG) and N3-methylthymine (3mT) on both single-stranded and double-stranded DNA.[11-13] AlkB repairs methylated DNA lesions by hydroxylation of the methyl group, coupled to the decarboxylation of αKG. The subsequent hydroxylated methyl group is proposed to decompose to formaldehyde and the unmethylated base (Figure III.1).

The presence of several AlkB human homologues, hABH1, hABH2, and hABH3, suggest that human cell lines employ a similar DNA repair strategy to that of *E. coli*.[1, 2] While the *E. coli* protein, AlkB, has broader substrate specificity, these human homologues are more specific in their activities and have distinct substrate preferences. Recently, hABH1 has been shown to demethylate 3mC in DNA as well as RNA.[14] Both hABH2 and hABH3 repair 1mA and 3mC, but hABH2 prefers double-stranded DNA while AlkB and hABH3 prefer single-stranded DNA.[1] Additionally, AlkB and hABH3 also repair RNA lesions.[1]

In addition to the methylated DNA lesions that AlkB repairs, our collaborator in Professor John Essigmann's group discovered that AlkB also repairs 1,N^6-ethenoadenine (εA) and 3,N^4-ethenocytosine (εC), while it reduces the toxicity and mutagenicity of 1,N^6-

ethanoadenine (EA).[4, 5] These lesions are a result of oxidative stress, or vinyl chloride and its metabolites. A recent study reports that hABH2 also repairs εA on dsDNA.[15] Another study showed that hABH3 also repairs εA, though much less efficiently than AlkB, and it is unclear if the activity is real or a result of a contaminant.[16] It is proposed that AlkB repairs εA by performing an epoxidation reaction on the ethenolated lesions rather than hydroxylation as in the case of methylated lesions. The epoxide ring on the ethenolated substrate is thought to open by hydrolysis to yield glycol, and the two-carbon adduct leaves as the dialdehyde glyoxal (Figure III.2).[4] Although AlkB does not completely repair EA, it reduces the toxicity and mutagenicity of EA via oxidative dealkylation as well.[5] It is proposed that AlkB hydroxylates the C8 carbon of EA, and the ethano-ring opens to form an acetaldehyde covalently attached to the N9 of adenine (Figure III.2), which could be further metabolized.[5]

Recently, the crystal structures of AlkB with and without a trimer DNA substrate (TmAT) were solved, showing AlkB with an overall structural fold that belongs to the cupin superfamily.[3] Enzymes in this superfamily catalyze a range of diverse reactions but share a conserved core structure, composed of a β-barrel or "jelly roll" motif.[17] The active site of these enzymes has a conserved structural motif that binds a divalent iron center through three protein ligands, two histidines and one carboxylate from an aspartate residue, which together form the facial triad.[3] These three conserved residues coordinate the iron center on one face of the octahedral site, and the other three sites are coordinated by water when the enzyme is in the resting state.

A mechanism for AlkB has been proposed based on structural and other biophysical studies (Figure III.3).[3, 18-20] The catalytic cycle begins by binding αKG to the iron in a bidentate fashion, replacing two water molecules in the iron center, and the binding of substrate displaces the last water to give a five coordinate iron center, which primes the site for binding dioyxgen

(shown as one step in Figure III.3). The decarboxylation of αKG promotes the cleavage of the O-O bond, thus concomitantly forming the oxo-ferryl species. This high valent iron intermediate is believed to be responsible for hydrogen abstraction and subsequently hydroxylates the methyl group on the damaged base. According to the crystal structure, the methyl group on the DNA lesion is too far away for hydrogen abstraction to occur without oxo-ferryl migration (not shown in figure).[3] Lastly, formaldehyde, undamaged base, and succinate leave the enzyme active site with three waters coming in to complete the iron hydration, as in the beginning of the catalytic cycle. The proposed mechanism does not discriminate between AlkB reactions with methylated substrates versus ethenolated substrates. The published structures have revealed interactions between the enzyme and its methylated substrate. However, there is no structural information on AlkB with ethenolated DNA lesions. It remain unclear how AlkB accommodates the bulkier ethenolated substrate in its active site or how AlkB repairs different alkylated substrates by different mechanisms.

This chapter describes the new purification scheme for AlkB, hABH2 and hABH3. The purified enzyme had been used for biochemical studies by Professor John Essigmann's laboratory. The crystallization and structural analysis of the substrate-free form of AlkBΔN11 is discussed.

III.C. Materials and Methods

The original genes for all proteins (AlkB, hABH2 and hABH3) described in this Chapter were a generous gift from Professor Leona Samson's laboratory. These genes were all in pET28a vectors. Construct reengineering described below was made in Professor Catherine Drennan's laboratory. The original vector for AlkB had a frame shift, which was corrected by sub-cloning

the gene. The AlkBΔN11 construct was made by sub-cloning. The hABH2 vector had an Arg203 to Cys mutation, which was corrected by site-directed mutagenesis. The hABH2-MBP fusion construct was made by sub-cloning. The hABH3 vector was fine. The All the DNA substrates used in crystallization were synthesized by Jim Delaney in Professor John Essigmann's laboratory.

Subcloning, Expression and Purification of AlkB and AlkBΔN11

The AlkB gene was subcloned using the forward primer AlkB-NdeI-F 5'-GGAATTC<u>CATATG</u>TTGGATCTGTTTGCCGATGCTGAACCG (underlined sequence is the Nde I restriction enzyme recognition site) and the reverse primer AlkB-EcoRI-R 5'-CG<u>GAATTC</u>TTATTCTTTTTTACCTGCCTGACGG (underlined sequence is the EcoR I restriction enzyme recognition site). The PCR amplification used an annealing temperature of 55°C and a 2 min extension time over 30 cycles. The amplified gene was digested with Nde I and EcoR I (New England Biolabs) and subcloned into a pET28a vector (Novagen) that had also been digested with Nde I and EcoR I. The sequence of the AlkB-pET28a clone was confirmed by DNA sequence analysis and was transformed into *E. coli* Rosetta 2(DE3) cells for expression.

To create the AlkBΔN11 construct (the first 11 residues at the N-terminus were removed from the full length AlkB gene) construct, the AlkB gene was subcloned using the forward primer AlkBΔN11-NdeI-F 5'-GGAATTC<u>CATATG</u>CAAGAGCCACTGGCGGCTGGTGCGG (underlined sequence is the Nde I restriction enzyme recognition site, and the beginning of the gene starts at the twelfth amino acid) and the reverse primer AlkBΔN11-XhoI-R 5'-CCG<u>CTCGAG</u>GCCTTGAAAATATAGGTTTTCTTTTTTACCTGCCTG (underlined sequence is the Xho I restriction enzyme recognition site, and highlighted sequence in red is the Tobacco

126

Etch Virus (TEV) recognition site). The PCR amplified gene was digested with Nde I and Xho I (New England Biolabs) and subcloned into a pET24a vector (Novagen) that was also digested with Nde I and Xho I. The sequence of the AlkBΔN11-pET24a clone was confirmed by DNA sequence analysis and was transformed into *E. coli* BL21(DE3) cells for expression.

Overnight cultures were grown in 250 mL of Luria-Bertani (LB) media with kanamycin (50 μg/mL) at 37°C with constant agitation at 250 rpm. The overnight culture was used to inoculate 4 L of LB media with kanamycin antibiotics. Cells were allowed to grow at 37°C until optical density at 600 nm reached 0.4 − 0.8, at which point isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM for induction. At this point, the cultures were transferred to 30°C for an additional 4 hr. Cells were harvested by centrifugation at 6000 x g for 20 min, and cell pellets were frozen in liquid nitrogen and stored at -80°C.

Cells were homogenized by sonication in lysis buffer (10 mM Tris, pH 7.34, 300 mM NaCl, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 5% glycerol, 1 mM β-mercaptoethanol (BME)). Cell debris was removed by centrifugation at 35,000 x g for 30 min. The supernatant was loaded onto a Ni-NTA column (Qiagen) that was pre-equilibrated with lysis buffer; 1 mL of resin slurry to 1 g of wet cell paste was use to pack the Ni-NTA column. The column was washed with 10 and 20 column volumes (CV) of lysis buffer with 10 mM and 20 mM imidazole, respectively. The His-tagged AlkB protein was eluted in 5 CV of lysis buffer with 70 mM and 250 mM imidazole. It was important to use two concentrations of imidazole (70 mM and 250 mM) to elute the protein from the Ni-NTA column; otherwise, the protein precipitated as it came off the column. The protein that eluted from the Ni-NTA column was pooled, mixed with thrombin (0.005U/10 μg protein), and dialyzed against thrombin cleavage buffer (20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$, 1 mM BME) overnight at 4°C. AlkB protein was then concentrated,

exchanged into 50 mM 2-[(tris(hydroxymethyl)methyl)amino]-1-ethanesulfonic acid (TES), pH 7.1 buffer, and loaded onto the SP Sepharose cation exchange column (Amersham Biosciences) that had been equilibrated with the same buffer. AlkB was eluted with a linear gradient of NaCl (0.0 – 1.0 M), and fractions containing AlkB were pooled, concentrated, and flash frozen for storage at -80°C. The final purified protein was analyzed by SDS-PAGE electrophoresis (Figure III.6). AlkBΔN11 protein was purified using the same procedure, with the exception that the His-tag was removed using TEV instead of thrombin.

Initial AlkB Crystallization, and Co-crystallization of AlkBΔN11 with εA Substrates

Initial crystallization screens for AlkB crystals yield "promising looking" phase separation by mixing 1 μL of AlkB (in 50 mM TES, pH 7.1) and 1 μL of precipitant solution (0.1 M HEPES, pH 7.3 and 10% w/v PEG 6000) (Hampton Research) in a drop placed over 0.5 mL of precipitant solution. Further optimization with N,N-dimethyldecylamine-β-oxide (DDAO) detergent (Hampton Research) dramatically improved the phase separation into needle haystack clusters (Figure III.4). These haystack crystals are not diffraction quality crystals; they are thin and non-single. At the time of further optimization for these crystals, Hunt *et. al* solved the crystal structure of AlkΔN11 with and without TmAT substrate.[3] Therefore, optimization of this crystal condition was not pursued further, as AlkBΔN11 crystals were reproducible with the published conditions.[3]

AlkBΔN11 crystals were obtained using the published condition with the hanging drop vapor diffusion method under anaerobic conditions.[3] AlkBΔN11 (10 mg/mL) in 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 1 mM BME were mixed with 0.2 M sodium formate, 10% glycerol, and 22% w/v PEG 3350 precipitant solution in a 1:1 ratio. αKG (Sigma), $Fe(NH_4)_2(SO_4)_2$ and

TεAT DNA substrate were added to the protein stock solution to a final concentration of 5 mM, 2 mM and 1.8 mM, respectively, prior to the crystallization setup. Large crystals appeared after 2 – 3 days (Figure III.5) Crystallization of AlkBΔN11 with 5'-phosphorylated trimer substrate (pTεAT) and etheno-adenine nucleotide (εA) were set up in the same manner as described above. No cryoprotectant was used prior to cryo-cooling using liquid nitrogen.

Data Collection and Processing of AlkBΔN11 Crystals Co-crystallized With TεAT, pTεAT, and εA Nucleotide Substrates

Datasets for the AlkBΔN11 crystals co-crystallized with TεAT and εA nucleotide substrates were collected to 1.2 and 1.5 Å resolution, respectively, at the Advanced Photon Source (Argonne, Illinois) on beamline 24ID-C Northeast Collaborative Access Team. A dataset for the AlkBΔN11 crystal co-crystallized with pTεAT substrate was collected to 1.8 Å resolution at the home source on a Rigaku system. All AlkBΔN11 crystals (with TεAT, pTεAT, and εA nucleotide substrates) obtained in this study belong to the P1 space group, and contained similar cell dimensions (40.1 Å, 40.9 Å, 73.7 Å, 74.7°, 77.0°, 63.2°). The space group and cell dimension of the AlkBΔN11 crystals obtained here were different from the space group and cell dimension of the published AlkBΔN11 with TmAT substrate. All datasets were integrated and scaled using Denzo and Scalepack, respectively.[21] Data collection statistics are shown in Table III.1.

AlkBΔN11 Structure Determination and Refinement

The structure of AlkBΔN11 co-crystallized with TεAT was solved by molecular replacement using PHASER[22] to 2.5 Å resolution. The structure of AlkB (2FDJ)[3], solved by

Hunts *et. al,* without any ligands or water molecules was used as a search model. The best solution had rotational and translational correlation coefficients of 16.2 and 10.5, respectively, with two AlkBΔN11 molecules per asymmetric unit. The PHASER output solution was refined by rigid body, followed by minimization and simulated annealing in CNS.[23] Phases were extended to 1.2 Å resolution, and waters were added to the model after a few rounds of refinement. Iterative rounds of model building and refinement were done in COOT and Refmac5.[24, 25] No non-crystallographic symmetry (NCS) restraints or sigma cutoff were used during the refinement.

Since the AlkBΔN11 crystals co-crystallized with pTεAT and εA nucleotide substrates were isomorphous with the AlkBΔN11 crystal co-crystallized with TεAT, a rigid body refinement followed by simulated annealing refinement was used to obtain a model. Further model building and refinements were done in COOT and Refmac5, respectively.[24, 25] No NCS restraints or sigma cutoff were used during the refinement. Composite omit maps calculated using CNS[23] simulated annealing were used to validate all the models. Refinement statistics for all models are shown in Table III.2.

The final structure of AlkBΔN11 co-crystallized with TεAT substrate was refined to 1.2 Å resolution with R_{cryst} of 19.8% and R_{free} of 21.4%. The final structure of AlkBΔN11 co-crystallized with pTεAT sbustrate was refined to 1.8 Å resolution with R_{cryst} of 23.1% and R_{free} of 26.7%, with the same reflection test set as for the AlkBΔN11-TεAT structure. The final structure of AlkBΔN11 co-crystallized with εA nucleotide substrate was refined to 1.5 Å resolution with R_{cryst} of 22.7% and R_{free} of 24.8%, with the same reflection test set as all the other structures. For all of the structures, no substrate was modeled due to little or no electron density observed where DNA is expected to bind.

Methyl Methanesulfonate (MMS) Plate Assay For hABH2 Activity

The MMS plate was made with two layers of 2% LB agar. A generic square cell culture dish plate was used, 13 mL of 2% LB agar with 0 – 0.06% of MMS gradient was poured at the bottom layer of the plate. After the bottom layer had dried, another layer of 25 mL of 2% LB agar with no MMS was poured on top of the plate. The control plate has no MMS drug in either layer of LB agar.

AlkB, hABH2, and hABH3 in pET28a vector, and empty pET28a vector were transformed into HK82 (AlkB deficient) *E. coli* cells. Each of these cells was grown in 5 mL of LB media with kanamycin (50 µg/mL) at 37°C overnight. The cells were spun down gently and resuspended in 1 mL of LB media. Prior to stamping the cells onto the MMS plates, 1 mL of cells was mixed with 1 mL of 1% LB agar. The long edge of a specimen glass slide was used for stamping the cells on the MMS plate.

Expression and Purification of hABH2 by Maltose-Binding Protein Fusion

hABH2 was initially purified using a maltose-binding protein (MBP) fusion in an attempt to avoid the protein forming inclusion bodies. The hABH2 gene was subcloned into the pMAL vector that contained a MBP fusion. The hABH2-pMAL vector was transformed into Rosetta 2(DE3)pLysS cells (Novagen). An overnight culture of hABH2 in Rosetta 2(DE3)pLysS cells was used to inoculate 4 L of LB media supplemented with glucose (2 g/L) and ampicillin (20 µg/mL). Cells were agitated at 37°C and 250 rpm until the optical density at 600 nm reached 0.5, at which point IPTG was added to a final concentration of 1mM for induction. Cells were allowed to continue to grow at 25°C overnight. Cells were harvested by centrifugation at 6000 x g for 20 min, and cell pellets were frozen in liquid nitrogen and stored at -80°C.

Cells were homogenized by sonication in lysis buffer (20 mM Tris-HCL, pH 7.4, 200 mM NaCl, 1 mM EDTA, and 10 mM BME) with 1 mg/mL lysozyme (Sigma). Cell debris was removed by centrifugation at 35,000 x g for 30 min. The supernatant was loaded onto an amylose column (New England Biolabs) that was pre-equilibrated with lysis buffer. The column was washed with 20 column volumes (CV) of lysis buffer. The hABH2-MBP fusion protein was eluted in 5 CV of lysis buffer with 10 mM maltose, and dialyzed overnight at 4°C against lysis buffer with 0.2 mg of Enterokinase (New England Biolab) to cleave the MBP. The protein was then loaded onto an amylose column (New England Biolab) that had been equilibrated with lysis buffer to remove the cleaved MBP and any hABH2-MBP fusion protein that had not yet been cleaved. The flow through from the amylase column containing hABH2 was then concentrated, and loaded onto a S75 size exclusion column (Amersham Biosciences) that had been equilibrated with storage buffer (50 mM TAPS, pH 8.0, 300 mM NaCl, 10% glycerol, and 1 mM BME). After the lengthy hABH2 purification, there was no detectable soluble hABH2 protein in the final FPLC elute. This method for hABH2 purification was thus abandoned.

Expression and Purification of hABH2 by His-tag

The expression conditions for hABH2 had been optimized by adjusting the induction temperature and the time length of the growth culture post-induction. Maximum expression was achieved at 37°C, with hABH2 allowed to grow for 4-6 hrs post-induction.

An overnight culture of hABH2 in Rosetta 2(DE3)pLysS cells was used to inoculate 4 L of LB media with kanamycin (50 μg/mL). Cells were agitated at 37°C and 250 rpm until the optical density at 600 nm reached 0.5, at which point IPTG was added to a final concentration of 1mM for induction. Cells were allowed to continue to grow at 37°C for an additional 4 – 6 hr.

Cells were harvested by centrifugation at 6000 x g for 20 min, and cell pellets were frozen in liquid nitrogen and stored at -80°C.

Cells were homogenized by sonication in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 1 mM BME) with 1 mg/mL lysozyme (Sigma). Cell debris was removed by centrifugation at 35,000 x g for 30 min. The supernatant was loaded onto a Ni-NTA column (Qiagen) that was pre-equilibrated with lysis buffer. The column was washed with 20 column volumes (CV) of lysis buffer with 20 mM imidazole. The His-tagged hABH2 was eluted in 5 CV of lysis buffer with 70 mM and 250 mM imidazole. It was important to use two concentrations of imidazole (70 mM and 250 mM) to elute the protein from the Ni-NTA column; otherwise, the protein precipitated as it came off the column. The eluted hABH2 was dialyzed against 50 mM NaH_2PO_4, pH 8.0 and 1 mM BME overnight in 4°C. The protein was then loaded onto the SP Sepharose cation exchange column (Amersham Biosciences) that had been equilibrated with dialysis buffer, and hABH2 was eluted with a linear gradient of NaCl (0.0 – 1.0 M). hABH2 was directly collected for *in vitro* activity assay, performed by Jim Delaney in John Essigmann's group. SDS-PAGE electrophoresis was used to analyze purity (Figure III.8).

Expression and Purification of hABH3

Overnight culture of his-tagged hABH3 in Rosetta 2(DE3)pLysS cells was used to inoculate 4 L of LB media with kanamycin (50 μg/mL). Cells were agitated at 37°C and 250 rpm until the optical density at 600 nm reached 0.6 – 0.8, at which point IPTG was added to a final concentration of 1 mM for induction. Cells were allowed to continue to grow at 30°C for an additional 6 hr. Cells were harvested by centrifugation at 6000 x g for 20 min, and cell pellets were frozen in liquid nitrogen and stored at -80°C.

Cells were homogenized by sonication in lysis buffer (25 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1 mM EDTA and 1 mM BME) with 1 mg/mL lysozyme (Sigma). Cell debris was removed by centrifugation at 35,000 x g for 30 min. The supernatant was loaded onto a Ni-NTA column (Qiagen) that was pre-equilibrated with lysis buffer. The column was washed with 20 CV of lysis buffer with 20 mM imidazole. The His-tagged hABH3 was eluted in 5 CV of lysis buffer with 70 mM and 250 mM imidazole. It was important to use two concentrations of imidazole (70 mM and 250 mM) to elute the protein from the Ni-NTA column; otherwise, the protein precipitated as it came off the column. To remove the His-tag, the eluted hABH3 was mixed with thrombin (0.005U/10 μg protein), and the mixture was dialyzed against thrombin cleavage buffer (20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$, 1 mM BME) for 4 hr at room temperature. The thrombin digested protein was eluted through another Ni-NTA column where the uncleaved His-tag protein separated. The protein was then concentrated, and loaded onto the S75 size exclusion column (Amersham Biosciences) that had been equilibrated with storage buffer (50 mM TAPS, pH 8.0, 300 mM NaCl, 10% glycerol, and 1 mM BME). The fractions with hABH3 were pooled, and loaded onto the SP Sepharose cation exchange column (Amersham Biosciences) that had been equilibrated with storage buffer, and the protein was eluted with a linear gradient of NaCl (0.0 – 1.0 M). hABH3 was directly collected for *in vitro* activity assay, performed by Jim Delaney in John Essigmann's group. SDS-PAGE electrophoresis confirmed the hABH3 was >95% pure (Figure III.6).

III.D. Results

MMS Plate Assay of AlkB, hABH1-3

Since the *in vitro* assay using purified hABH2 was not initially reproducible by following the assay conditions published,[1] the MMS plate assay was used to confirm that our constructs of AlkB and its human homologues, hABH2-3, were functioning properly *in vivo*. An empty pET28a vector was used as a negative control, the hABH2 clone from Leona Samson's lab that has been shown to be active was used as a positive control, and AlkB and hABH2-3 genes were tested in the AlkB-deficient *E. coli* (HK82) cells. When these cells were plated on a no-drug control plate, all of them survived and grew across the entire plate (Figure III.7.a). When these cells were plated on a 0% - 0.06% (from left to right) MMS plate, all cells grew to a lesser extent compared to their growth on the control plate (Figure III.7.b). In row A, the empty pET28a vector encodes no DNA repair enzyme, and thus the cells died at MMS concentration greater than about 0.03%. In row B, the positive control grew more than the negative control in the presence of MMS, indicating functional protein was produced to sustain cell survival at a greater concentration of MMS. In rows C-E, vectors containing the AlkB, hABH2 and hABH3 gene from our lab were shown to grow similarly compared to the positive control in row B. Therefore, this data suggested that the clones of AlkB and hABH2-3 were correctly constructed, and they encode functional proteins.

Expression and Purification of AlkB, AlkBΔN11 and hABH3

For crystallization and biochemical studies, AlkB and its homologue proteins were purified with His-tags on a Ni-NTA chromatography column in the initial step of purification. The His-tag was cleaved in the next step of the purification, either by thrombin for AlkB and

hABH3, or by TEV for AlkBΔN11. Further purification of the proteins was performed by either the use of a SP Sepharose cation exchange fast-protein liquid chromatography (FPLC) (for AlkB and AlkBΔN11), or by size exclusion FPLC (for hABH3). For hABH3, an extra step of purification employing the SP Sepharose cation exchange FPLC was used. Purifying these proteins with a SP Sepharose cation exchange column was important for removing the exonuclease activity from the protein. Expression of an 8 L culture of the AlkB, AlkBΔN11, and hABH3 produced about 5 mL of 10 mg/mL of purified protein, which was a sufficient quantity for biochemical and crystallization studies. Protein concentrations were determined by absorbance at 280 nm using the theoretical extinction coefficients of each protein; 32,680 $M^{-1}cm^{-1}$ for AlkB, 31,065 $M^{-1}cm^{-1}$ for hABH2, and 66,140 $M^{-1}cm^{-1}$ for hABH3.

All the purified proteins, AlkB (Figure III.6.a), AlkBΔN11 (Figure III.6.b) and hABH3 (Figure III.6.c), were >95% pure as judged by SDS-PAGE electrophoresis. All the purified proteins were also free of exonuclease activity; this is very important, as exonuclease degrades the DNA substrate that is required for the biochemical assay. The activities of AlkB and its homologues were determined by Jim Delaney in John Essigmann's lab using mass spectrometry. All purified AlkB, AlkBΔN11 and hABH3 samples were fully active (data not shown).

Optimized Expression, Purification, and Activity of hABH2

The previous lack of hABH2 *in vitro* activity was believed to be due to the low yield in protein purification, a result of hABH2 being expressed in an inclusion body. To avoid the low yields and insolubility issues, many expression conditions and constructs have been tested. Expression at low temperature (15°C overnight after induction) produced soluble hABH2 protein; however, the yield was very low. The exact concentration was unable to be determined using UV

absorbance at 280 nm, but the presence of hABH2 could be detected by FPLC. Expression of hABH2 fused to a maltose-binding protein (MBP) at higher expression temperatures (30°C for 4-6 hr after induction) produced soluble hABH2-MBP protein; however, hABH2 became unstable and precipitated upon cleaving the MBP. Finally, the expression condition that yielded a reasonable amount of soluble hABH2 protein was using very high temperature (37°C) for a short induction time (4 hr or less). Although the expression of hABH2 at 37°C for a short period of time yielded a reasonable amount of soluble protein, it was still significantly less than the yields of AlkB or hABH3 (an 8 L culture yielded only 3-4 mL of < 1 mg/mL of hABH2). In addition, the purification procedure for hABH2 has not been optimized to obtain the quality and quantity of hABH2 required for crystallization.

The SDS-PAGE electrophoresis of the purified hABH2 showed a mixture of full length hABH2 and its degraded fragments in solution (Figure III.8). Full length hABH2 has been reported to be labile and easily degraded.[1] There were three major fragments resulting from hABH2 proteolysis: the molecular weights of each as judged by SDS-PAGE electrophoresis is approximately 27 kDa, 17 kDa and 12 kDa. Full length hABH2 and three degraded fragments were present at equal amounts in solution as judged by the SDS-PAGE electrophoresis gel. In spite of the proteolysis of hABH2, the purified hABH2 mixture was active. An *in vitro* activity assay done by Jim Delaney showed that hABH2 repaired about 60% of the methylated 16-mer dsDNA in 30 min after incubation with the mixture of full length and degraded hABH2 fragments (Figure III.9). In Figure III.9, the envelope of peaks to the left corresponds to the complementary strand of the 16-mer double-stranded DNA. The middle envelope of peaks corresponds to the repaired DNA strand. The envelope of peaks on the right corresponds to the methylated DNA strand that was left unrepaired by hABH2.

The ethenoadenine lesion was recently found to be a new type of substrate repaired by AlkB.[4] Studies have shown that AlkB repairs εA lesions *in vivo*, as well as the TεAT trimer *in vitro*.[4, 16] However, there is no kinetic data present for the repair of TεAT trimer *in vitro*; a study has only shown that complete repair of TεAT trimer occurs in 1 h with a 10-fold excess of substrate over the enzyme.[16] The structure of AlkBΔN11 with αKG and iron was determined at 1.2 Å resolution. The structure belongs to the P1 space group with cell dimension 40.1 Å, 40.9 Å, 73.7 Å, 74.7°, 77.0°, 63.2°. There are two molecules of AlkBΔN11 in the asymmetric unit (ASU), related by two-fold non-crystallographic symmetry (Figure III.10). The structure exhibits the same jelly-roll motif as all the other enzymes in the αKG mononuclear iron-dependent superfamily. The active site contains iron, coordinated by His131, Asp133, His187, αKG (bidentate binding), and a water molecule. This structure is very similar to both the published structures with (2FD8)[3] and without (2FDJ)[3] TmAT substrate bound, and with RMSDs of 0.911 Å over 197 C_α atoms and 0.873 Å over 174 C_α atoms, respectively.

While the active site of AlkBΔN11 is well assembled and the iron-coordinating ligands are well ordered (the average B-factor for His131, Asp133, His187 and αKG is 10 $Å^2$), the substrate DNA is disordered, as indicated by the lack of electron density near the DNA binding site (Figure III.11). In an attempt to improve the electron density for the DNA substrate, the TεAT trimer was modeled and refined against the data; however, this technique failed to produce a better quality electron density map for the DNA substrate. Structural alignment between this AlkBΔN11 structure and the published AlkB structure with TmAT substrate bound (2FD8)[3] predicts that the DNA binding sites of each monomer in the ASU point toward each other in our structure. The reason why TεAT does not bind becomes apparent when two TεAT molecules are

modeled into each monomer of AlkBΔN11 in the ASU. The space between the two monomers is not large enough to accommodate two TεAT molecules, as the ends of the DNA would clash if two DNA molecules were present in the ASU (Figure III.12). Although the DNA substrate was pre-mixed with the protein and the mixture was allowed to equilibrate at room temperature for 15-30 min prior to crystallization setup, the DNA substrate either did not bind to the protein or unbound from the protein during the time that it took for crystal formation. Since this crystal packing clearly does not favor substrate binding, as there is not enough room between the protein molecules to accommodate two pieces of trimer DNA, the crystal selectively forms with the AlkBΔN11 protein that does not have substrate DNA bound to it.

It is possible that AlkBΔN11 did not crystallize without DNA bound because the TεAT binding was not tight enough to form a stable complex between AlkBΔN11 and the TεAT substrate for the duration of 2-3 days of crystal growth. In an attempt to capture a more stable AlkB-DNA complex, a 5′-phosphorylated TεAT (pTεAT) substrate was used in hope that the 5′ phosphate will create multiple hydrogen bonds with the protein and stabilize the protein-substrate interactions.

Co-crystallization of AlkBΔN11 with pTεAT

The structure of AlkBΔN11, αKG and iron co-crystallized with pTεAT was determined at 1.8 Å resolution. Like the structure of AlkBΔN11 co-crystallized with TεAT, this structure belongs to the P1 space group with cell dimension 39.3 Å, 39.7Å, 72.3 Å, 74.6°, 77.0°, 63.5°. The cell dimension of this crystal is similar to the the AlkBΔN11 structure co-crystallized with TεAT. There are two molecules of AlkBΔN11 in the ASU, and they are related by two-fold non-crystallographic symmetry. This AlkBΔN11 structure has RMSDs of 0.910 Å over 197 C_α atoms

and 0.829 Å over 174 C_α atoms with the published AlkB structures with (2FD8)[3] and without (2FDJ)[3] DNA bound, respectively.

Since this crystal form has the same space group and cell dimension as the crystal of AlkBΔN11 co-crystallized with TεAT, we were concerned that crystal packing in this crystal would be the same, thus preventing pTεAT from binding to AlkBΔN11. The structures of AlkBΔN11 co-crystallized with TεAT and pTεAT are very similar, with an RMSD of 0.382 $Å^2$ over 199 C_α atoms. Upon several rounds of model building and refinement, an electron density map was made without DNA substrate in the model. The $2F_o$-F_c map shows little electron density where DNA binds, suggesting either pTεAT is not present or is disordered with low occupancies at the substrate binding site (Figure III.13).

Since the addition of an extra phosphate group at the 5′ end of the DNA trimer does not increase its binding to the protein, variations of the length of DNA substrate could be tested. A longer piece of DNA could provide more protein-substrate interactions to stabilize substrate binding in the crystal, or a shorter piece of DNA may allow the substrate to fit between the AlkBΔN11 molecules in the existing crystal form. A structure of AlkBΔN11 with a longer piece of DNA substrate would also be more physiologically relevant. However, there is no data about the optimal length for DNA binding to AlkB, and a larger piece of DNA might alter the crystallization condition, such that sparse matrix screening of the protein-substrate complex would be required. On the other hand, if a smaller substrate, εA nucleotide, is used, the DNA substrate might be able to soak into the AlkBΔN11 crystals, eliminating the issue of low DNA binding affinity to AlkB. In addition, previous studies have shown that AlkB has repair activity for a lone nucleotide lesion, e.g. 1-methyl adenosine nucleotide.[26, 27] With these advantages and

disadvantages in mind, we decided to pursue AlkBΔN11 co-crystallization with the εA nucleotide.

Co-crystallization of AlkBΔN11 with εA Nucleotide

Nucleotide substrates have been shown to be repaired by AlkB, though with less efficiency than longer DNA substrates, with a k_{cat} of ~4 min^{-1} and 7.5 min^{-1} for nucleotide and trimer, respectively.[27] However, AlkB's reactivity to εA nucleotide has not been tested.

The structure of AlkBΔN11, αKG and iron co-crystallized with εA nucleotide was determined at 1.5 Å resolution. Like the other structures of AlkBΔN11 co-crystallized with a trimer DNA substrate, this structure also belongs to the P1 space group with cell dimension 39.2 Å, 39.7 Å, 72.3 Å, 74.9°, 77.2°, 63.6°. The cell dimension of this crystal is similar to the AlkBΔN11 structure co-crystallized with TεAT and pTεAT. There are two molecules of AlkBΔN11 in the ASU that are related by two-fold non-crystallographic symmetry. Like the other two AlkBΔN11 structures that are co-crystallized with TεAT and pTεAT, this structure has RMSDs of 0.891 Å over 197 C_α atoms and 0.835 Å over 180 C_α atoms for the published AlkBΔN11 structures with (2FD8)[3] and without (2FDJ)[3] TmAT substrate bound, respectively. The three AlkBΔN11 structures mentioned here are remarkably similar to one another (Table III.3).

After several rounds of model building and refinement without modeled substrate, a $2F_o$-F_c map was made to see if there is unbiased electron density for the substrate. The $2F_o$-F_c map shows a blob of electron density where the etheno lesion should bind, which at initial structural analysis is a good sign of having the εA nucleotide bound to the protein (Figure III.14). However, there is no electron density for the sugar or the phosphate of the nucleotide. An εA nucleotide

was modeled and refined against the data; however, the small blob of density where the ethenoadenine ring should be located was not enough electron density to refine the rest of the εA nucleotide in the active site. Previous studies have shown that the apparent K_m between a 1-methyl adenine trimer DNA substrate and AlkB is about 3 μM.[27] It is possible that the ethenoadenine substrate has weaker binding compared to 1-methyl adenine. In addition, the published crystal structure of AlkBΔN11 with T-1-methyl-A-T substrate reveals that most of the protein-DNA interactions are from the DNA backbone, and the methylated nucleotide is only stabilized by the π stacking between Trp69 and the active site His131.[3] It is possible that with just the εA nucleotide, the π stacking between the aromatic residues are not tight enough for the εA nucleotide to stay in the AlkB active site during the crystallization experiment.

III.E. Discussion

AlkB is a DNA repair enzyme in *E. coli* with broad substrate specificities. It repairs one-carbon adduct lesions such as 1mA, 3mC, 1mG and 3mT on both DNA and RNA;[11-13] it also repairs two-carbon adduct lesions such as εA, εC and EA.[4, 5, 16] While these lesions are all repaired by AlkB in *E. coli*, only some are repaired by each human homologue of AlkB. For example, hABH2 repairs 1mA, 3mC, and εA on DNA only; and hABH3 repairs 1mA and 3mC on DNA and RNA. The differences between these enzyme active sites that account for their difference in substrate preference has been of interest to many scientists. In this work, AlkB, hABH2 and hABH3 were purified using a new purification method that allowed efficient removal of exonuclease activity. The important step for removing the exonuclease activity is running the protein through a cation exchange chromatography column.

Co-crystallization of AlkBΔN11 (purified by the new method described in this Chapter) with TεAT, pTεAT and εA nucleotide yielded a substrate-free form of the AlkBΔN11 structure. The most obvious difference between the published AlkBΔN11 structure with TmAT substrate bound and the structures obtained here is the difference among the space groups, P4$_3$ versus P1, respectively (Table III.4). In space group P4$_3$, there is a screw axis down the long side of the unit cell. A screw axis is a combination of rotational and translational elements. For example, the screw axis in P4$_3$ space group is a combination of rotating the red molecule around the axis by 360°/4 or 90° (Figure III.15.a) and then translating it by ¾ of the unit cell length (Figure III.15.b) to end up in the green molecule. P1 is the simplest space group, and it has no rotational or screw axes. Any asymmetric unit in the P1 space group is related to another molecule by a simple translation along one of its unit cell axes (Figure III.15.c). Although the differences in space group alone cannot explain why no TεAT or εA nucleotide substrates were observed in these crystals, the difference in space group results in different crystal packing, which may be one reason why no substrate is bound in our P1 cell. One of the published AlkBΔN11 structure without DNA substrate also crystallized in a P1 space group (Table III.4); it is possible the crystal packing in a P1 cell does not have space for the DNA substrate. The variation in substrate used (TmAT versus TεAT) is unlikely to be responsible for the different quality of density observed in the active site.

AlkBΔN11 crystallized as a dimer in both the P4$_3$ and P1 space groups. However, the arrangement of the molecules in the dimer is different. In the case of the P4$_3$ space group, the two AlkB molecules were packed stacked on top of each other, such that one molecule's active site is pointing towards the bottom face of another molecule (Figure III.16). In the case of the P1 space group, the two molecules are pointing at each other, with one molecule's active site

pointing towards the active site of another molecule (Figure III.12). In Figure III.12, the substrates were modeled based on structural alignment with the published AlkBΔN11 structure with TmAT substrate bound; they were not modeled based on density from a dataset.

There was no substrate observed in the co-crystallization experiments of AlkBΔN11 with TεAT and pTεAT, the reason became clear after modeling the substrates in the P1 space group dimer. The P1 space group dimer has active sites of each monomer pointing towards each other due to the crystal packing, which would have resulted in the substrates of the two monomers clashing against each other if both of them were bound at once. Since the trimer substrates cannot be bound to both molecules at the same time, each monomer has either no substrate bound or one substrate per dimer. The sharing of substrate within the dimer may lead to the substrate "swinging" between the two monomers, resulting in a substrate occupancy of ~50% of the time for each monomer, which contributes to little or no density. Upon the realization of this setback to obtaining a structure of AlkBΔN11 with TεAT substrate bound, the first thought was to use a nucleotide substrate instead of the trimer DNA. Based on the structural alignment modeling of the substrate, a nucleotide lesion should be fully buried in the active site of AlkBΔN11, thereby resolving the issue of two substrates clashing against each other. However, we once again saw little or no substrate density in the εA co-crystallized form.

The published AlkBΔN11 structure with TmAT substrate bound and the structures obtained here are aligned to compare the differences in conformation that might explain why the substrate is not bound. The RMSD between the published AlkBΔN11 structure with TmAT bound and AlkBΔN11 co-crystallized with TεAT is 0.911 Å over 197 C_α atoms. When we examine the active site region of the aligned structures, one obvious difference is the conformation of the loop and the β-strand that follows, consisting of residues Gly74 to Ser79.

The published AlkBΔN11 with TmAT structure has a closed conformation where the loop and the β-strand are closed down at the substrate binding site to lock the substrate in place. In contrast, in the structure obtained here, without the TεAT bound, the loop and β-strand are farther from the substrate binding site, which is probably not the optimal conformation to lock down the substrate (Figure III.17). While this difference in conformation of the loop and the β-strand above the substrate binding site is obvious and apparent, it is unclear if this difference in conformations can by itself explain the lack of εA substrate binding.

Further study is necessary to elucidate the active site differences between AlkB with 1mA substrate and AlkB with εA substrate bound. A longer piece of DNA might be important for obtaining a structure of AlkBΔN11 with εA substrate bound. If a structure of AlkBΔN11 co-crystallized with εA nucleotide will be pursued in the future, testing AlkB's activity on εA nucleotide to ensure that εA nucleotide is also an AlkB substrate might also be important.

III.F. Acknowledgements

III.G. References

1. Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B., Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99, (26), 16660-16665.

2. Koivisto, P.; Duncan, T.; Trewick, S. C.; Lindahl, T.; Sedgwick, B., E. coli and human dioxygenases repair DNA adducts generated by simple alkylating agents. *Toxicology and Applied Pharmacology* **2004**, 197, (3), 201-201.

3. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F., Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **2006**, 439, (7078), 879-884.

4. Delaney, J. C.; Smeester, L.; Wong, C. Y.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M., AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nature Structural & Molecular Biology* **2005**, 12, (10), 855-860.

5. Frick, L. E.; Delaney, J. C.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Alleviation of 1,N-6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, (3), 755-760.

6. Jiricny, J., DNA repair: Bioinformatics helps reverse methylation damage. *Current Biology* **2002**, 12, (24), R846-R848.

7. Drablos, F.; Feyzi, E.; Aas, P. A.; Vaagbo, C. B.; Kavli, B.; Bratlie, M. S.; Pena-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E., Alkylation damage in DNA and RNA - repair mechanisms and medical significance. *DNA Repair* **2004**, 3, (11), 1389-1407.

8. Margison, G., A new damage limitation exercise: ironing (Fe(II) out minor DNA methylation lesions. *DNA Repair* **2002**, 1, (12), 1057-1061.

9. Sedgwick, B., Repairing DNA-methylation damage. *Nature Reviews Molecular Cell Biology* **2004**, 5, (2), 148-157.

10. Sedgwick, B.; Lindahl, T., Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* **2002**, 21, (58), 8886-8894.

11. Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B., Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002**, 419, (6903), 174-178.

12. Falnes, P. O.; Johansen, R. F.; Seeberg, E., AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **2002**, 419, (6903), 178-182.

13. Falnes, P. O., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Research* **2004**, 32, (21), 6260-6267.

14. Westbye, M. P.; Feyzi, E.; Aas, P. A.; Vagbo, C. B.; Talstad, V. A.; Kavli, B.; Hagen, L.; Sundheim, O.; Akbari, M.; Liabakk, N. B.; Slupphaug, G.; Otterlei, M.; Krokan, H. E., Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *Journal of Biological Chemistry* **2008**, 283, (36), 25046-25056.

15. Ringvoll, J.; Moen, M. N.; Nordstrand, L. M.; Meira, L. B.; Pang, B.; Bekkelund, A.; Dedon, P. C.; Bjelland, S.; Samson, L. D.; Falnes, P. O.; Klungland, A., AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Research* **2008**, 68, (11), 4142-4149.

16. Mishina, Y.; Yang, C. G.; He, C., Direct repair of the exocyclic DNA adduct 1,N-6-ethenoadenine by the DNA repair AlkB proteins. *Journal of the American Chemical Society* **2005**, 127, (42), 14594-14595.

17. Dunwell, J. M.; Culham, A.; Carter, C. E.; Sosa-Aguirre, C. R.; Goodenough, P. W., Evolution of functional diversity in the cupin superfamily. *Trends in Biochemical Sciences* **2001**, 26, (12), 740-746.

18. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M.; Krebs, C., The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: A high-spin Fe(IV) complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2003**, 42, (24), 7497-7508.

19. Price, J. C.; Barr, E. W.; Hoffart, L. M.; Krebs, C.; Bollinger, J. M., Kinetic dissection of the catalytic mechanism of taurine: alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2005**, 44, (22), 8138-8147.

20. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine :alpha-ketoglutarate dioxygenase (TauD). *Journal of the American Chemical Society* **2003**, 125, (43), 13008-13009.

21. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **1997**, 276, 307-326.

22. Read, R. J., Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D-Biological Crystallography* **2001**, 57, 1373-1382.

23. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **1998**, 54, 905-921.

24. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **2004,** 60, 2126-2132.

25. Vagin, A. A.; Steiner, R. A.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D-Biological Crystallography* **2004,** 60, 2184-2195.

26. Welford, R. W. D.; Schlemminger, I.; McNeill, L. A.; Hewitson, K. S.; Schofield, C. J., The selectivity and inhibition of AlkB. *Journal of Biological Chemistry* **2003,** 278, (12), 10157-10161.

27. Koivisto, P.; Duncan, T.; Lindahl, T.; Sedgwick, B., Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *Journal of Biological Chemistry* **2003,** 278, (45), 44348-44354.

III.H. Tables and Figures

Table III.1. Data collection and processing statistics for models of AlkBΔN11 co-crystallized with TεAT, pTεAT and εA nucleotide.

Structure	AlkBΔN11 Co-crystallized with TεAT	AlkBΔN11 Co-crystallized with pTεAT	AlkBΔN11 Co-crystallized with εA nucleotide
Crystal Parameter			
Space group	P1	P1	P1
Unit cell dimension a,b,c, (Å) α,β,γ (°)	40.1, 40.9, 73.7, 74.7, 77.0, 63.2	39.3, 39.7, 72.3, 74.6, 77.0, 63.5	39.2, 39.7, 72.3, 74.9, 77.2, 63.6
Data Collection			
Wavelength (Å)	1.0	1.0	1.0
Resolution limit[b] (Å)	50 – 1.20 (1.24 – 1.20)	30 – 1.86 (1.93 – 1.86)	50 – 1.50 (1.55 – 1.50)
R_{sym}[a,b] (%)	3.4 (19.9)	3.3 (12.3)	5.3 (24.7)
No. of unique reflections	113395	29625	44052
Redundancy[b]	3.6 (3.4)	3.3 (2.6)	2.1 (1.6)
Completeness[b] (%)	90.8 (90.4)	93.8 (86.2)	73.2 (44.9)
Mean I/sigma[b]	28.3 (6.7)	31.8 (6.3)	12.2 (2.9)

[a]$R_{sym} = (\Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|)/ \Sigma_{hkl}\Sigma_i I_i(hkl)$ for n independent reflections and i observations of a given reflection. $<I(hkl)>$ = average intensity of the ith observation.

[b]Numbers for the highest resolution shell are shown in parentheses.

Table III.2. Refinement statistics for models of AlkBΔN11 co-crystallized with TεAT, pTεAT and εA nucleotides.

Structure	AlkBΔN11 Co-crystallized with TεAT	AlkBΔN11 Co-crystallized with pTεAT	AlkBΔN11 Co-crystallized with εA nucleotide
$R_{cryst}/R_{free}{}^{a}$ (%)	19.8/21.4	23.1/26.7	22.7/24.8
Rmsd bond lengths (Å)	0.014	0.018	0.010
Rmsd bond angles (deg)	1.367	1.716	1.367
Ramachandran analysis			
Most favored (%)	88.3	90.7	90.9
Allowed (%)	10.8	8.4	8.5
Generously allowed (%)	0	0.3	0
Disallowed (%)	0.9	0.6	0.6

[a]$R_{cryst} = \Sigma_h ||F_o(h)| - |F_c(h)|| / \Sigma_h |F_o(h)|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is calculated the same way with a test set of reflections (10%) that are not used during refinement.

Table III.3. RMSD between different models of AlkBΔN11 co-crystallized with different

substrates.

Models	RMSD	Over # C_α Atoms
TεAT/pTεAT	0.382	199
TεAT/εA nucleotide	0.379	199
pTεAT/εA nucleotide	0.125	199

Table III.4. Space group comparison of the published structures with structures obtained here.

Published Structures[3]	Substrate	Space group	Cell Dimension a,b,c, (Å) α,β,γ (°)	Resolution (Å)
Co/αKG/TmAT	Yes	P4$_3$	41.3, 41.3, 116.8 90, 90, 90	2.10
Mn/αKG/TmAT	Yes	P4$_3$	40.7, 40.7, 118.3 90, 90, 90	2.10
Fe/αKG/TmAT	Yes	P4$_3$	40.7, 40.7, 118.3 90, 90, 90	2.30
Fe/SUC/TmAT	Yes	P4$_3$	40.5, 40.5, 117.4 90, 90, 90	2.20
Fe/SUC	No	P1	36.8, 38.7, 40.2 75.8, 75.0, 66.1	2.10
εA co-crystallization structures				
TεAT	No	P1	40.1, 40.9, 73.7, 74.7, 77.0, 63.2	1.20
pTεAT	No	P1	39.3, 39.7, 72.3, 74.6, 77.0, 63.5	1.86
εA nucleotide	No	P1	39.2, 39.7, 72.3, 74.9, 77.2, 63.6	1.50

Figure III.1. AlkB reactions with 1mA and 3mC substrates.

1mA

3mC

Fe(II)
AlkB

αKG
O_2

CO_2
succinate

CH_2O

adenine

cytosine
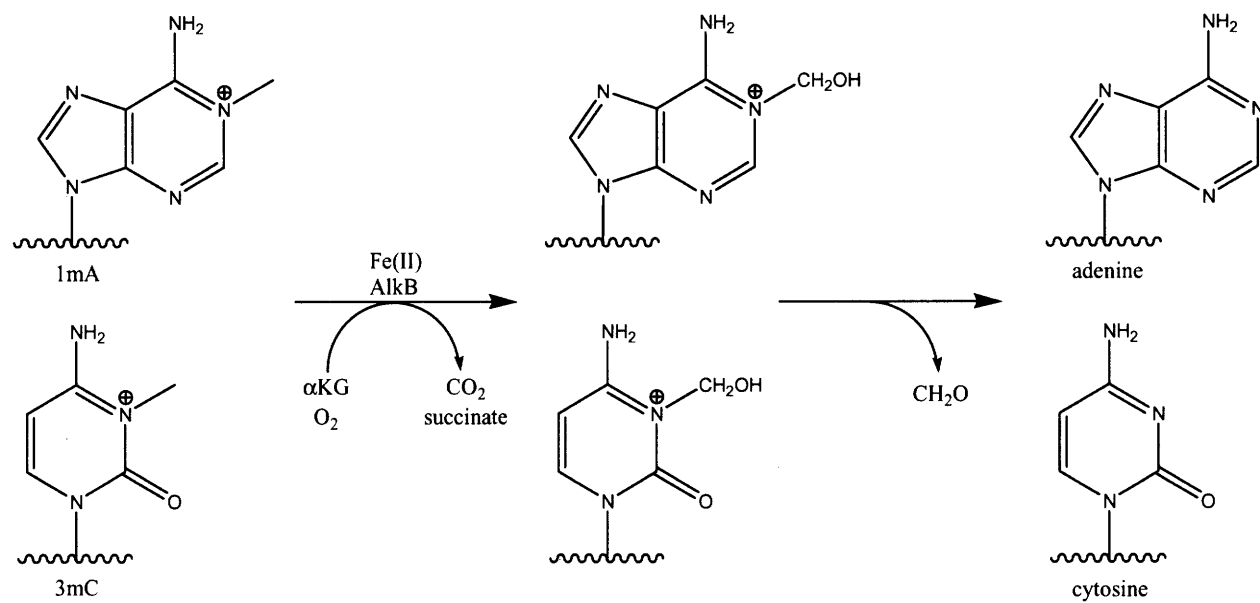
Figure III.2. Proposed reaction of AlkB with εA and EA substrates.[4, 5]

εA   εA epoxide   εA glycol   Adenine   Glyoxal

EA   εA epoxide   N^6-acetaldehyde
                                intermediate
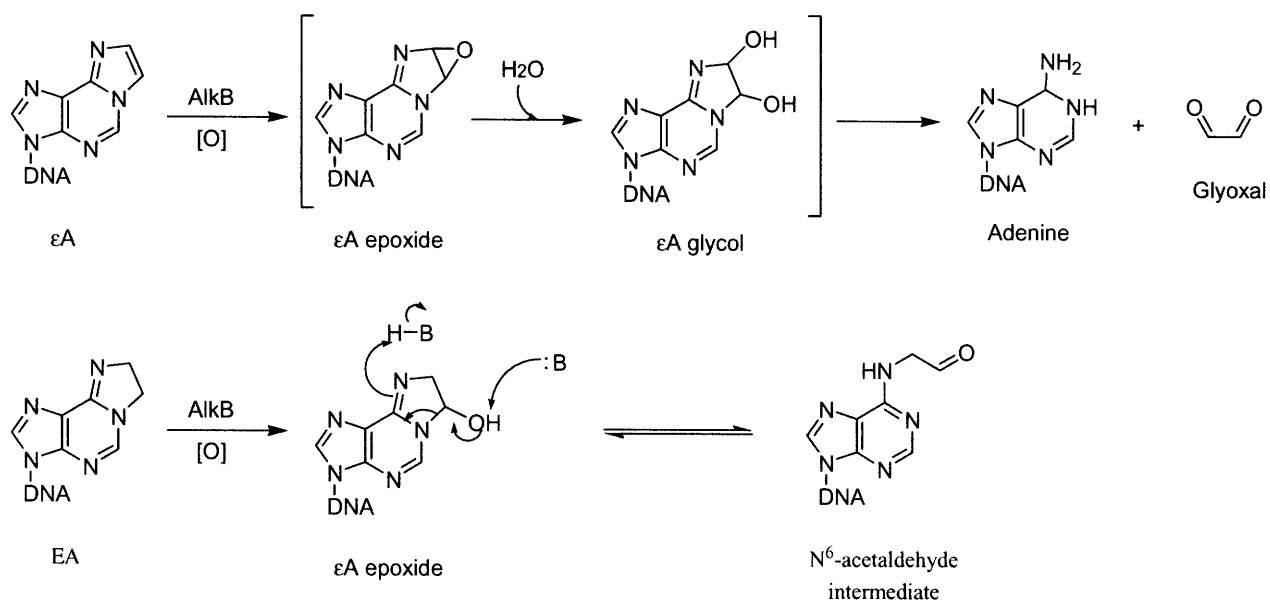
Figure III.3. Proposed mechanism of AlkB with methylated DNA. The oxo-ferryl migration is hypothesized (not shown in the figure) based on the crystal structure of AlkB with 1mA bound.[3]
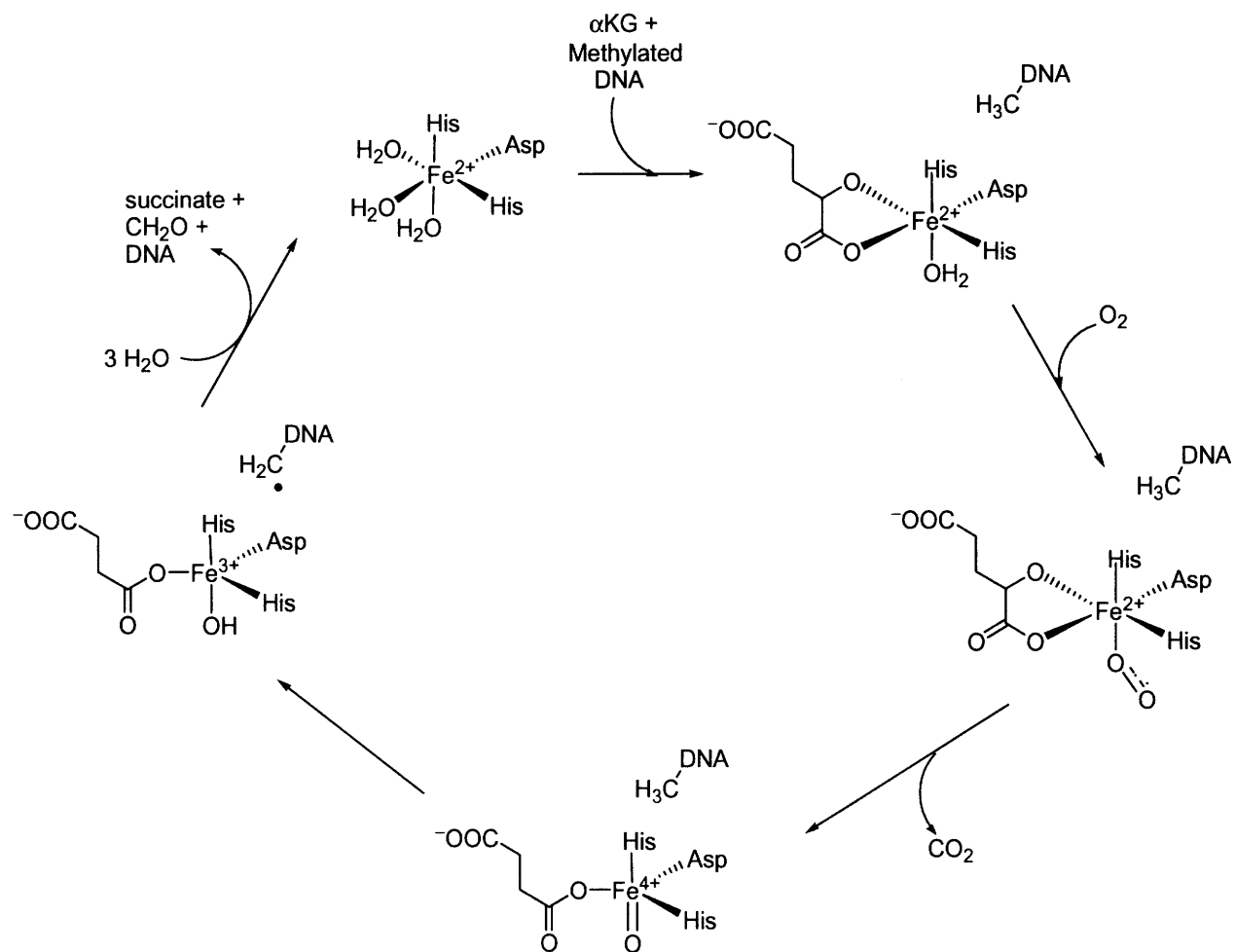
Figure III.4. Clusters of needle haystacks of AlkB grew from optimized conditions with N,N-dimethyldecylamine-β-oxide (DDAO) detergent.

**Figure III.5.** Crystal of AlkBΔN11 grown in the presence of TεAT substrate, shown under polarized light. The rystal grew in published conditions[3], but with a different unit cell.
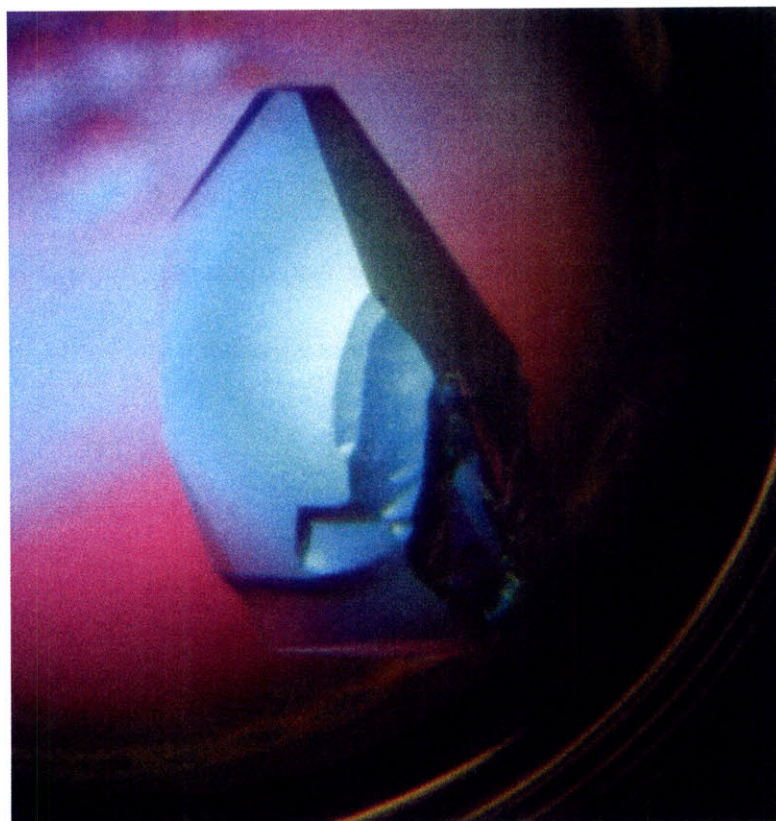
Figure III.6. SDS-PAGE electrophoresis gel of final purified AlkB, AlkBΔN11 and hABH3 proteins to be >95% purity. a) SDS-PAGE analysis of purified AlkB. Lanes 1 and 3 are the protein standard molecular weight marker. Lanes 2 and 4 are 50 μg and 80 μg of purified AlkB, respectively. The theoretical molecular weight of AlkB is 25 kDa. b) SDS-PAGE analysis of purified AlkBΔN11. Lane 1 is the protein standard molecular weight marker. Lane 2 is 80 μg of purified AlkBΔN11. The theoretical molecular weight of AlkBΔN11 is 24.7 kDa. c) SDS-PAGE analysis of purified hABH3. Lane 1 is the protein standard molecular weight marker. Lane 2 is 50 μg of purified hABH3. The theoretical molecular weight of hABH3 is 33 kDa.
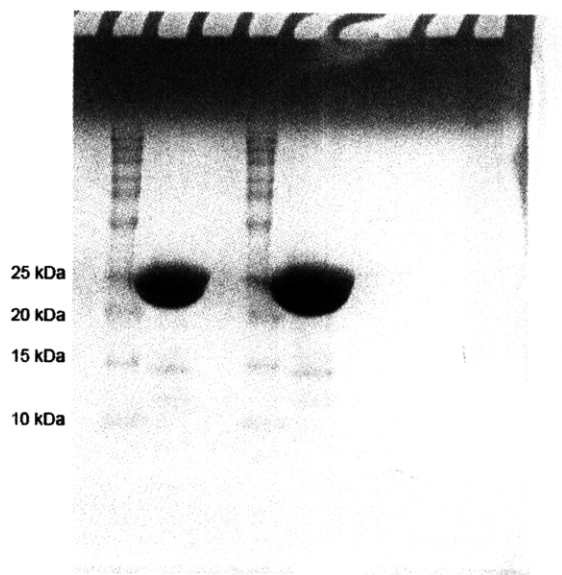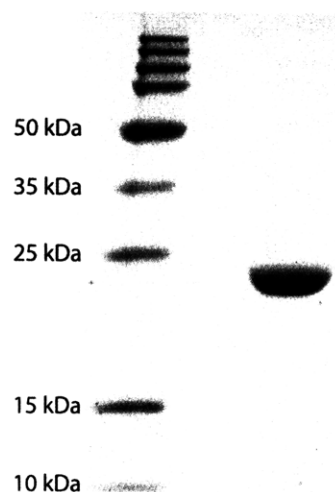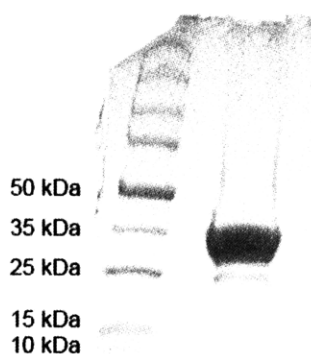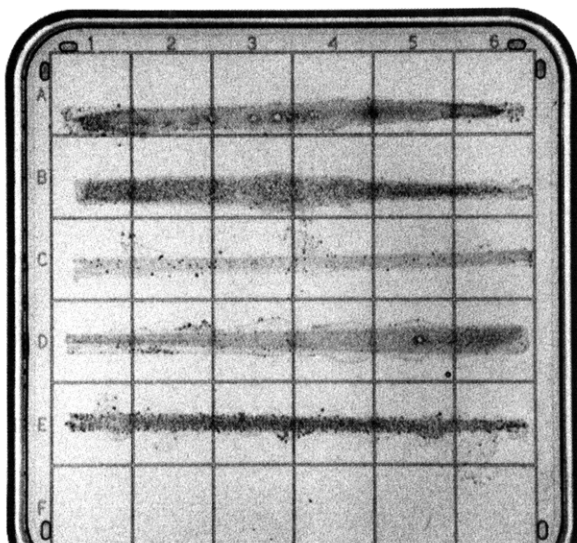
Figure III.7. MMS plate assay of AlkB, hABH2 and hABH3. Row A is the empty pET28a vector, as a negative control. Row B is the hABH2 from Professor Leona Samson's lab, as a positive control. Rows C-E are AlkB, hABH2 and hABH3 from our lab, respectively. a) Control plate with no MMS drug. b) Assay plate with MMS concentrations range from 0% (left) to 0.06% (right) on the plate.
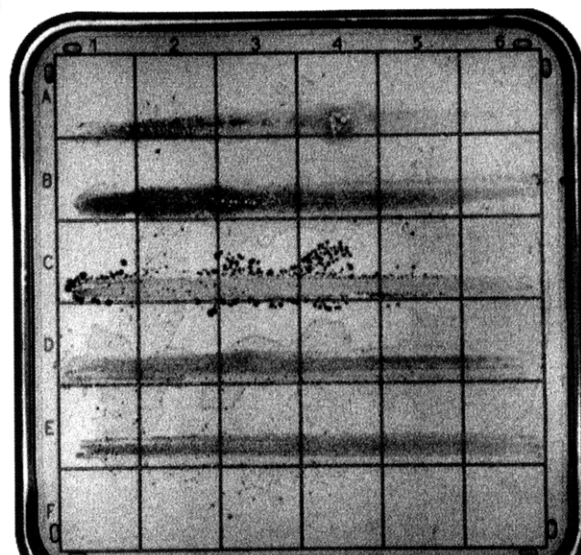
Figure III.8. SDS-PAGE electrophoresis gel of purified hABH2. Lane 1 is the protein standard molecular weight marker. Lane 2 is 50 μg of purified hABH2. The top band corresponds to full-length hABH2; theoretical molecular weight of full length hABH2 is 29.3 kDa. The lower three bands have molecular weights of approximately 27 kDa, 17 kDa, and 12 kDa.
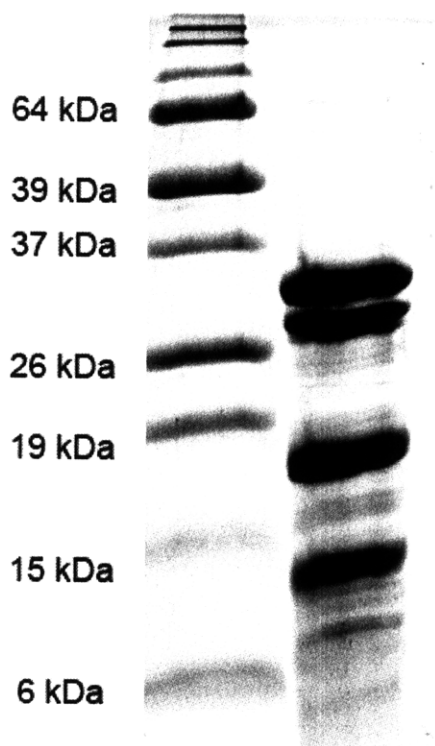
Figure III.9. MALDI-MS trace of hABH2 activity assay with 16-mer methylated double-stranded DNA. The envelope of peaks to the left corresponds to the complementary strand of the 16-mer double-stranded DNA. The middle envelope of peaks corresponds to the repaired DNA strand. The envelope of peaks on the right corresponds to the methylated DNA strand that was left unrepaired by hABH2.
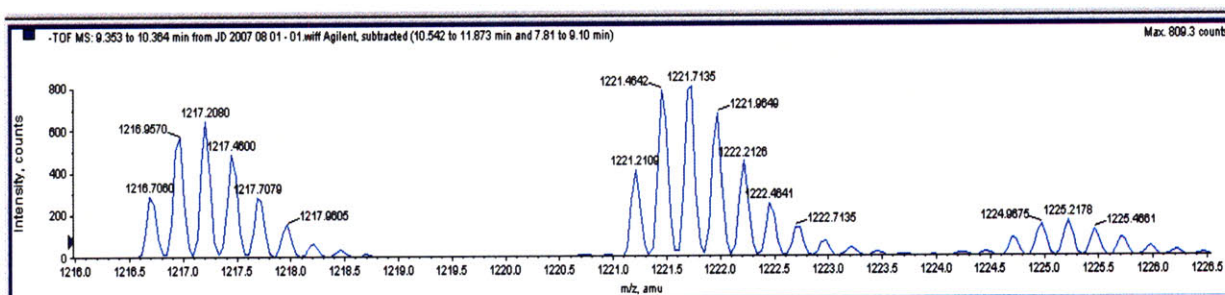
Figure III.10. Overall structure of AlkBΔN11 co-crystallized with TεAT. The crystallographic dimer is colored by molecule. The iron ligands: His131, Asp133, His187 and αKG are shown in stick representation. The iron metal center is shown in sphere representation. Substrate TεAT was not modeled.

Figure III.11. Active site of AlkBΔN11 co-crystallized with TεAT. The iron ligands: His131, Asp133, His187 and αKG are shown in stick representation. The iron metal center and water is shown in brown and red sphere representations, respectively. A $2F_o\text{-}F_c$ map contoured at 1σ is shown in blue mesh around the stick residues and the substrate binding region where the TεAT should be bound.

Figure III.12. AlkBΔN11 dimer with TmAT modeled in the P1 space group. The crystallographic dimer is colored by molecule, with the AlkBΔN11 protein molecules shown in cartoon representation. The ends of the DNA trimer substrates are shown in stick representation, with the clashing portion enlarged in the inset figure.
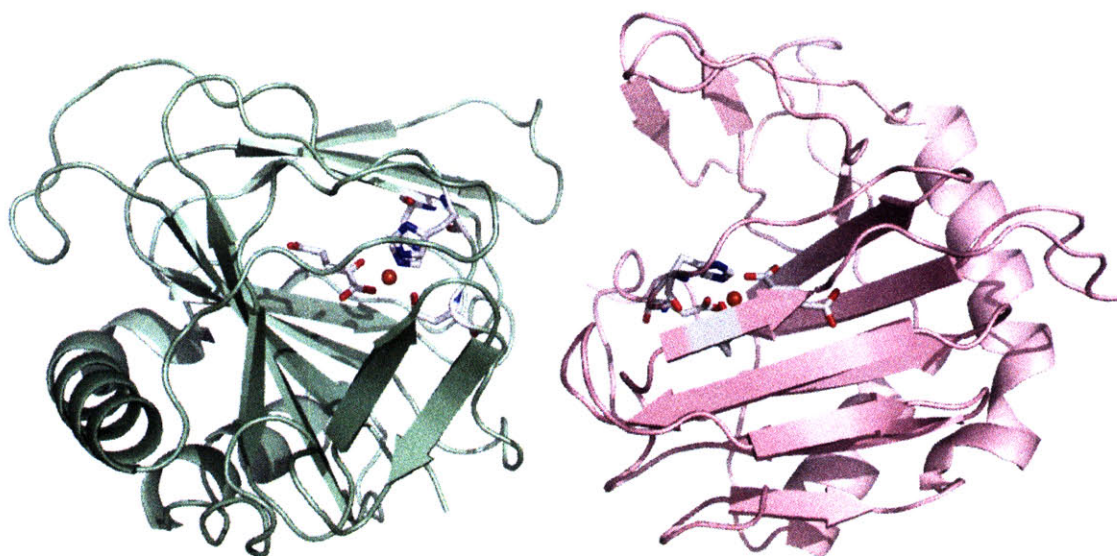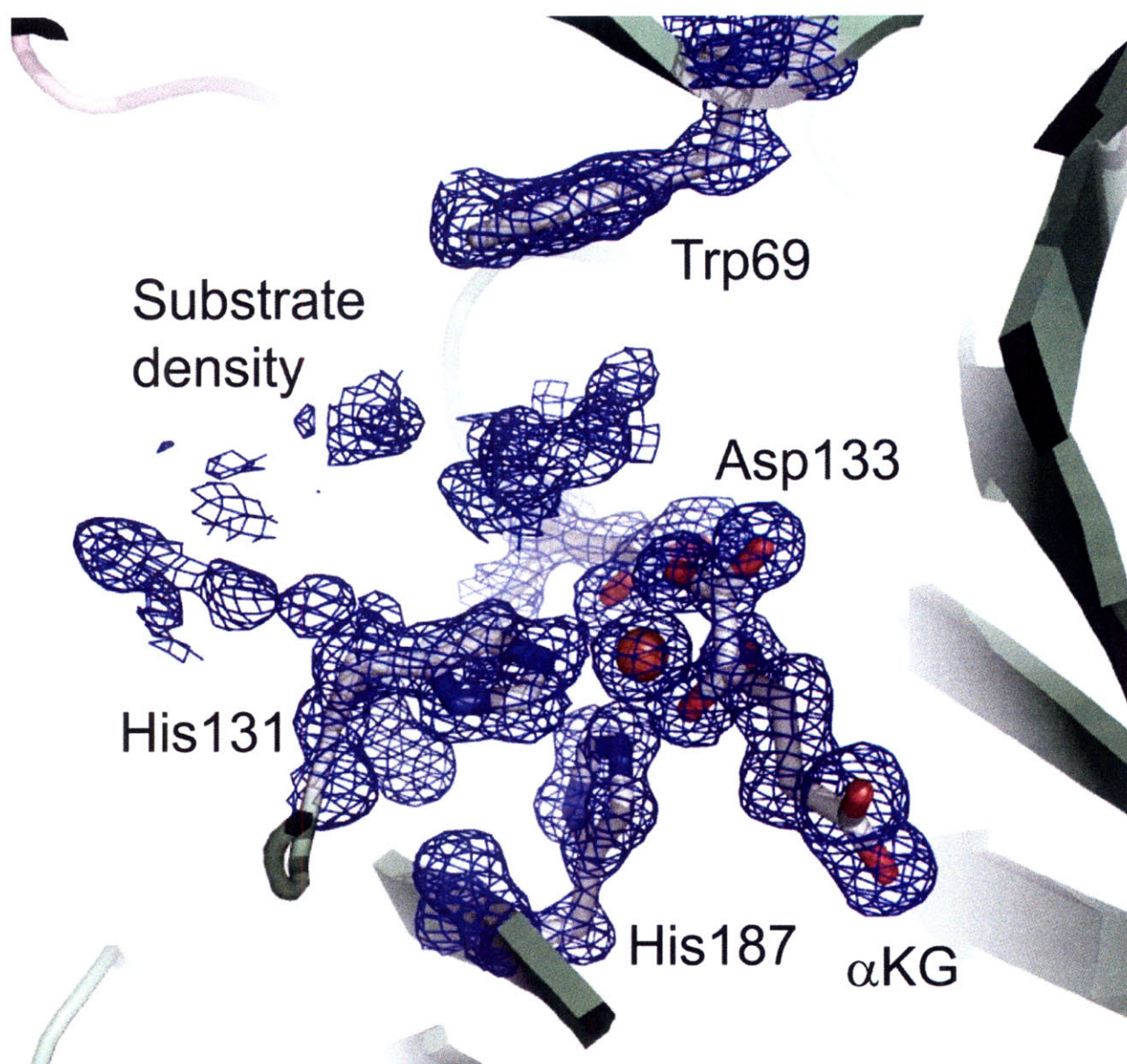
Figure III.13. Active site of AlkBΔN11 co-crystallized with pTεAT. The iron ligands: His131, Asp133, His187 and αKG are shown in stick representation. The iron metal center and water is shown in sphere representation. A $2F_o\text{-}F_c$ map contoured at 1σ is shown in blue mesh around the stick residues and the substrate binding region where the pTεAT should be bound.
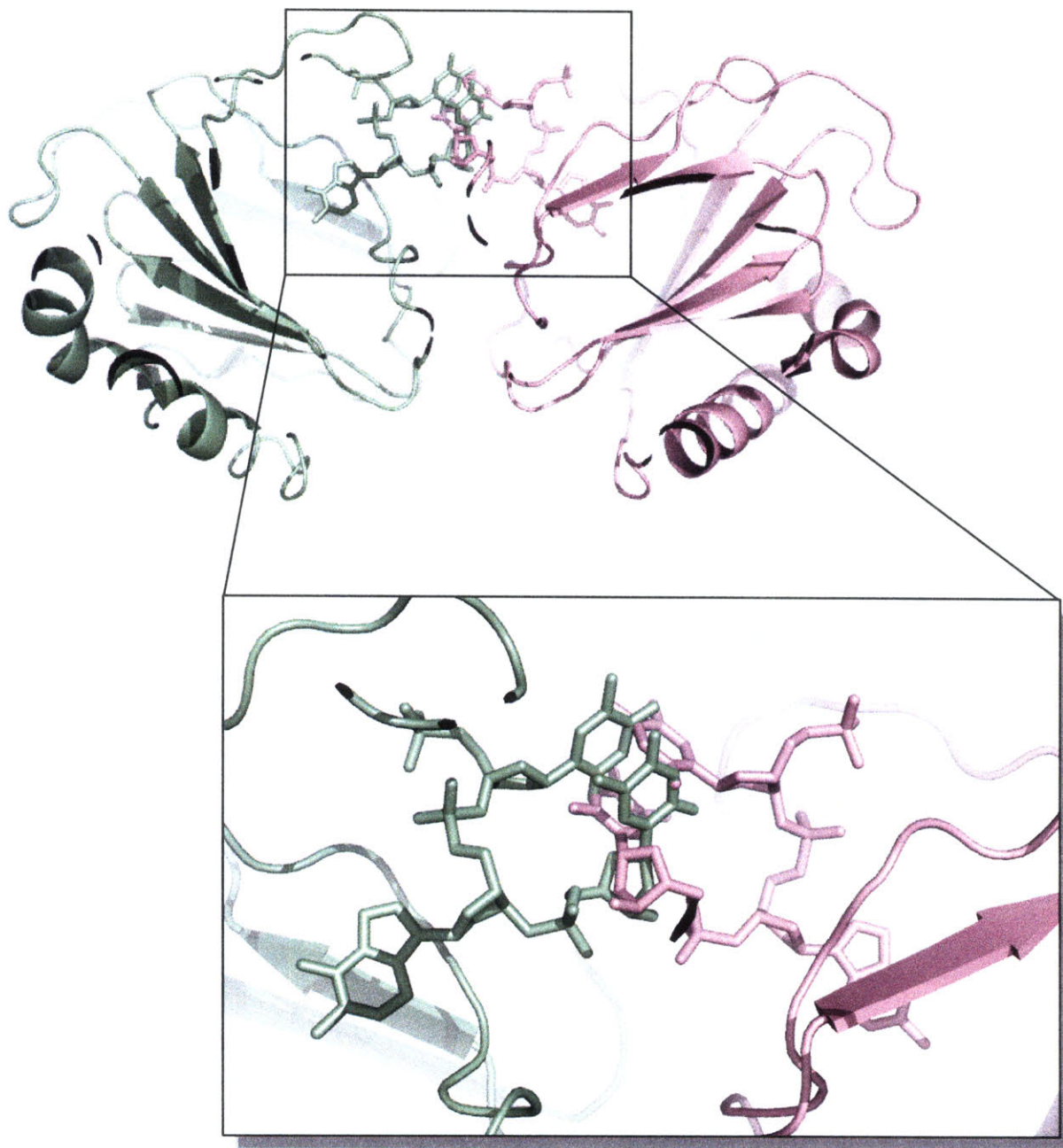
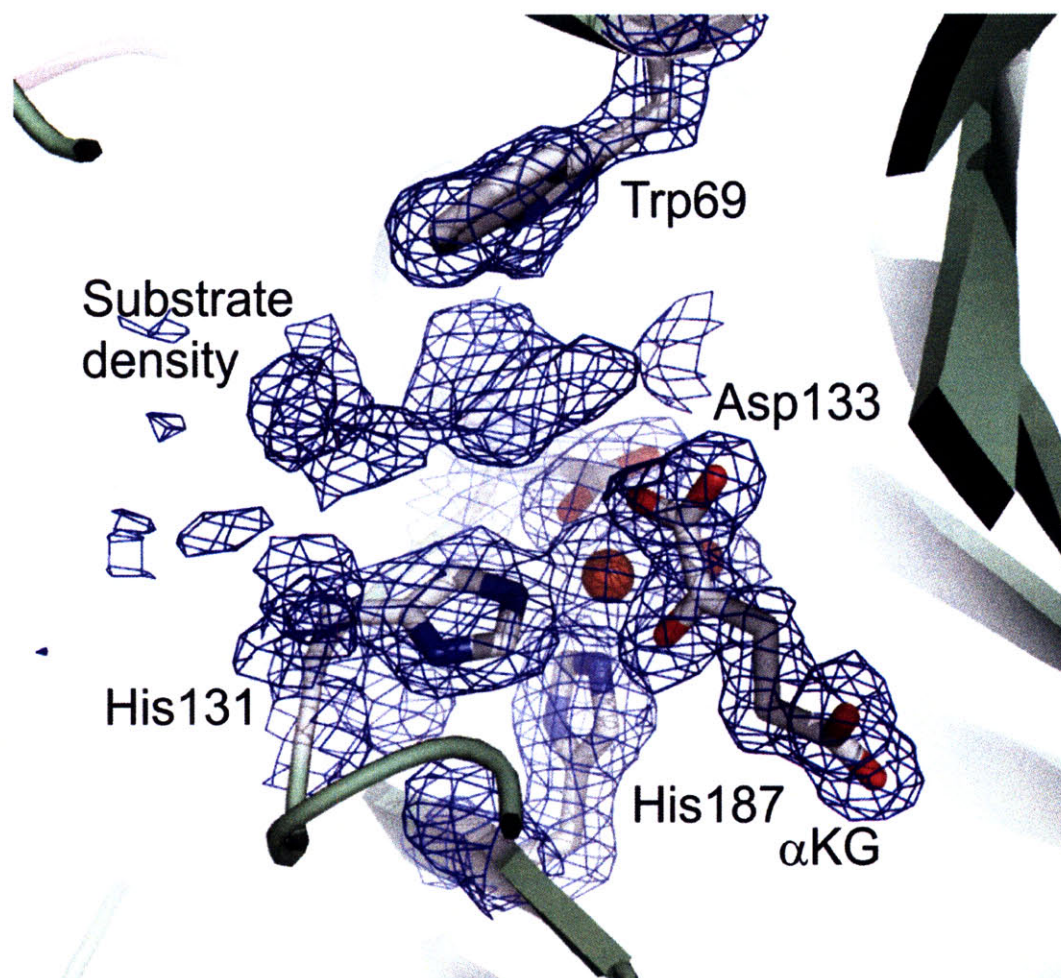**Figure III.14.** Active site of AlkBΔN11 co-crystallized with εA nucleotide. The iron ligands: His131, Asp133, His187 and αKG are shown in stick representation. The iron metal center and water is shown in sphere representation. A $2F_o$-F_c map contoured at 1σ is shown in blue mesh around the stick residues and the substrate binding region where the εA nucleotide should be bound.
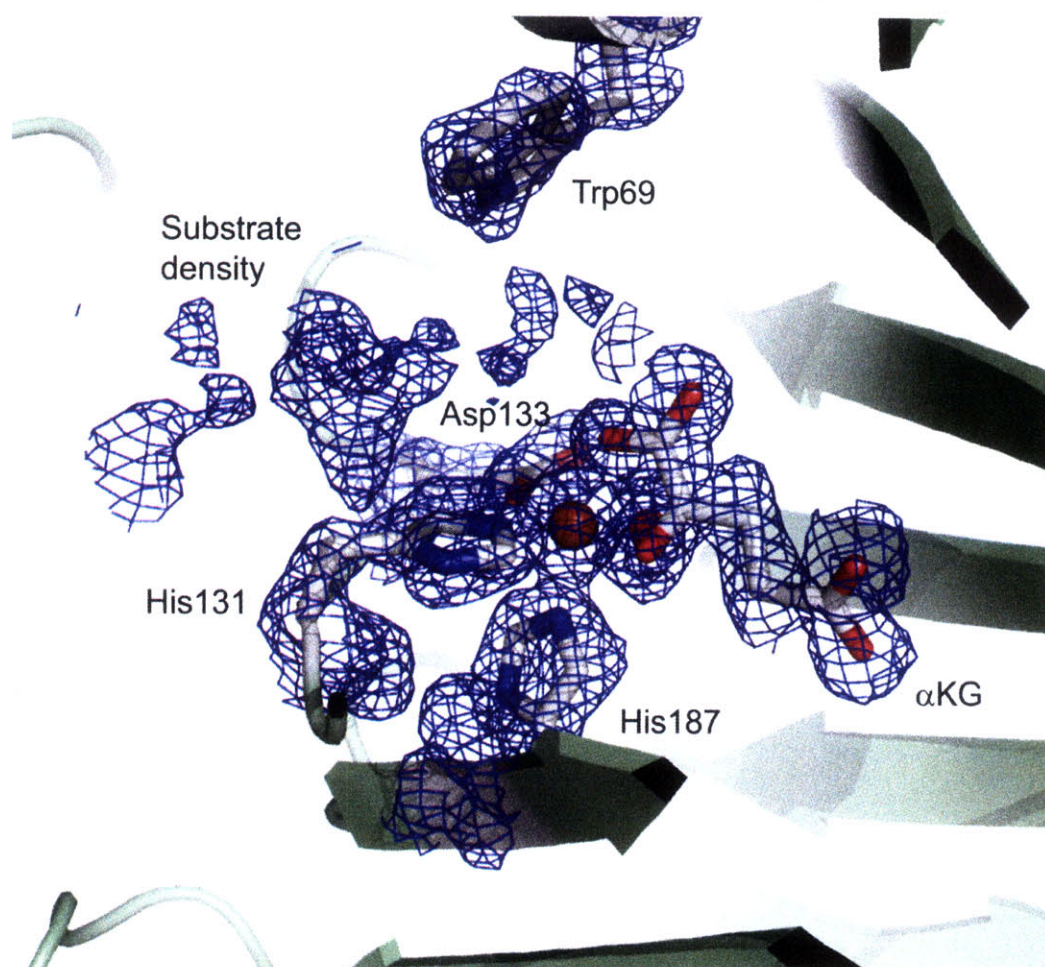
Figure III.15. Crystal packing in P4$_3$ and P1 space groups. a) Top view of P4$_3$ space group lattice packing. The unit cell is defined by the blue rectangular box. Four molecules of TmAT DNA substrate are shown in stick representation, and are colored by molecule. For clarity, the protein atoms have been omitted from the figure. A screw axis perpendicular to the paper is defined by a black circle. b) Side view of P4$_3$ space group lattice packing. The unit cell is defined by the blue rectangular box. Four molecules of TmAT DNA substrate are shown in stick representation, and are colored by molecule. For clarity, the protein atoms have been omitted from the figure. A screw axis is defined by a black line along the unit cell edge. An arrow indicates the direction of translation. c) Lattice packing of P1 space group. The unit cell is defined by the blue rectangular box. Two molecules of AlkBΔN11 are shown in cartoon representation and are colored by molecule. An arrow indicates the direction of translation.

Figure II.15.

a

Screw axis perpendicular to paper →

90 degree rotation

b

Screw axis

Translation by 3/4 of a unit cell

c

Translation by 1 unit cell

168
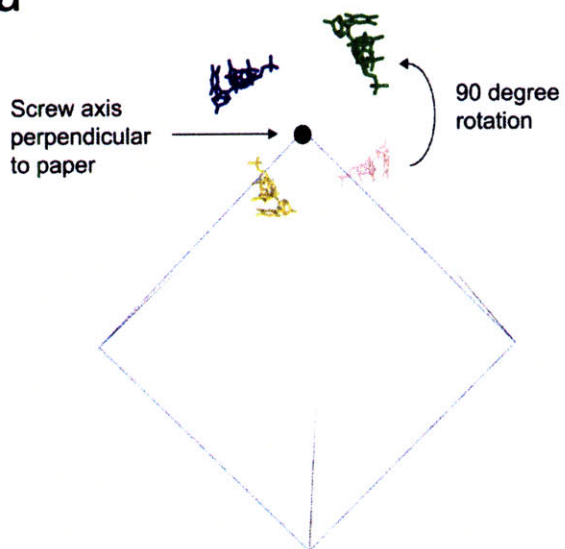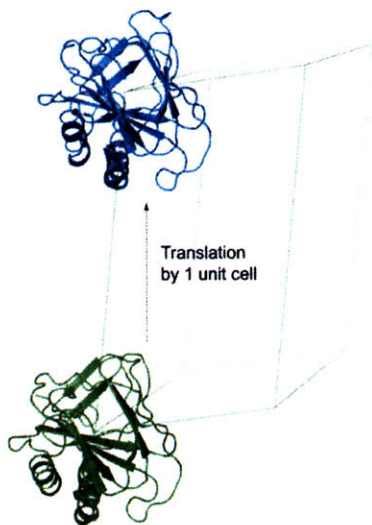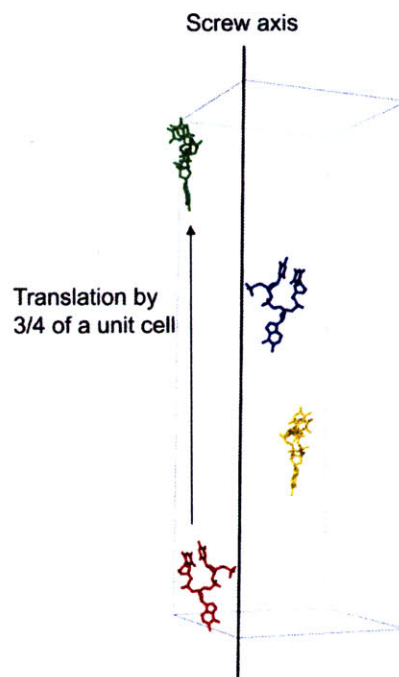
Figure III.16. AlkBΔN11 dimer with TmAT in P4$_3$ space group.[3] The crystallographic dimer is colored by molecule. The TmAT DNA substrates are shown in stick representation.
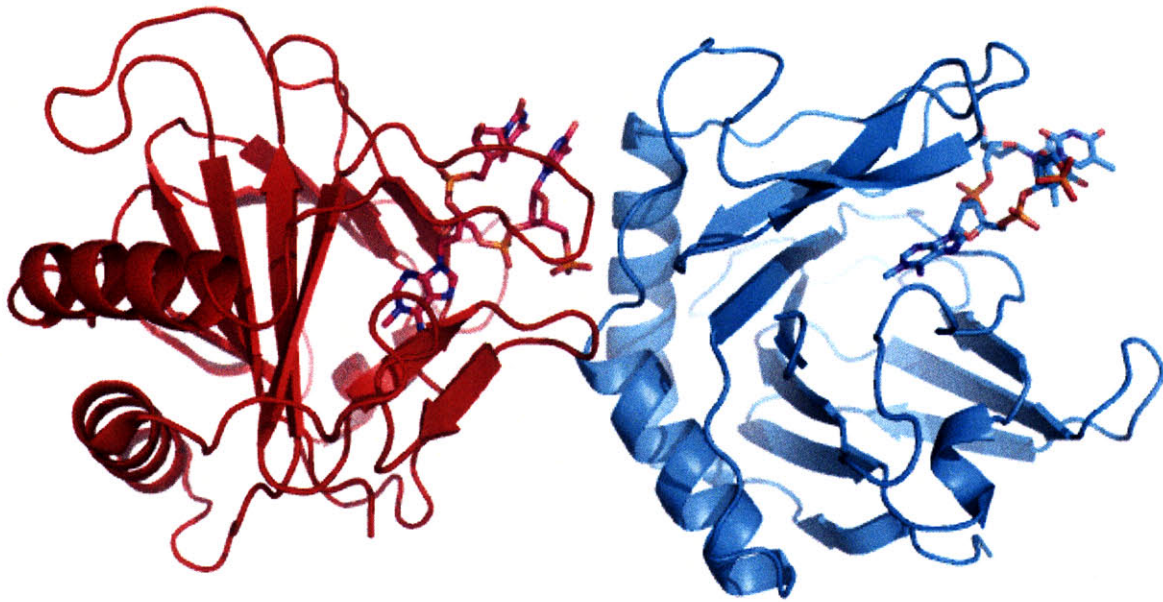
Figure III.17. Structural alignment of the published AlkBΔN11 with TmAT bound (orange) and the AlkBΔN11 without substrate (magenta) obtained when co-crystallized with TεAT. The His131, Trp69 and the DNA substrate are shown in stick representation. The rest of the protein is shown in cartoon representation. A black arrow is shown where the loop and the β-strand are in different conformations for the two structures.
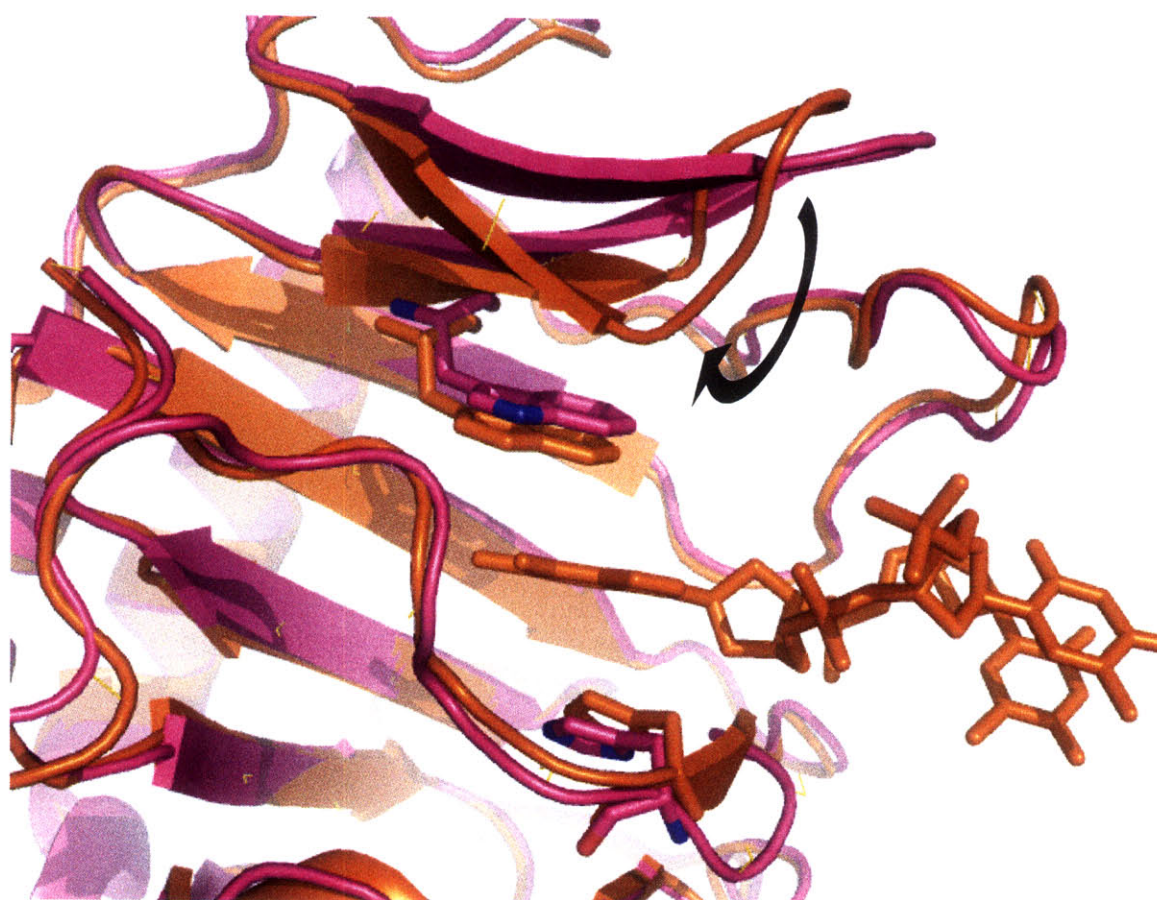
Chapter IV

Crystallographic Analysis of an Open Active Site Conformation of Non-heme Iron Halogenase CytC3

IV.A. Summary

CytC3, a member of the recently discovered class of non-heme Fe(II) and α-ketoglutarate (αKG)-dependent halogenases, catalyzes the double chlorination of L-2-aminobutyric acid (Aba) to produce a known *Streptomyces* antibiotic, γ,γ-dichloroaminobutyrate. Unlike the majority of the Fe(II)-αKG-dependent enzymes that catalyze hydroxylation reactions, halogenases catalyze a transfer of halides. To examine the important enzymatic features that discriminate between chlorination and hydroxylation, the crystal structures of CytC3 both with and without αKG/Fe(II) have been solved to 2.2 Å resolution. These structures capture CytC3 in an open active site conformation, in which no chloride is bound to iron. Comparison of the open conformation of CytC3 with the closed conformation of another non-heme iron halogenase, SyrB2, suggests two important criteria for creating an enzyme-bound Fe-Cl catalyst: (1) the presence of a hydrogen-bonding network between the chloride and surrounding residues, and (2) the presence of a hydrophobic pocket in which the chloride resides.

IV.B. Introduction

Many halogenated natural products have important therapeutic properties, such as anti-cancer and anti-fungal activities.[1] Examples of chlorinated natural products include salinisporamide A, a potent anti-cancer agent,[2] and syringomycin, an anti-fungal compound.[3] The incorporation of chlorine into natural products can be important for tailoring a compound's activity. For example, the addition of a single chlorine atom to the syringomycin scaffold increases its anti-fungal properties 4-fold.[4] The biosynthesis of natural products with cyclopropane ring structures can involve so called "cryptic halogenations", where a halide is used to functionalize a non-reactive group and is subsequently displaced by a substrate-derived enolate to generate the three member ring. Since chlorination by chemical synthesis can be environmentally unfriendly and technically challenging,[5] there is interest in enzymatic halogenation chemistry for the production of various natural products.

The biosynthesis of γ,γ-dichloroaminobutyrate in soil *Streptomyces* is carried out by a four component non-ribosomal peptide synthetase (NRPS) assembly, CytC1-CytC4.[6] CytC1 is a free-standing adenylation domain that loads the nonproteinogenic amino acid L-2-aminobutyric acid (Aba) onto the phosphopantetheine arm of a thiolation domain, CytC2 (Figure IV.1). While the amino acid is tethered to CytC2, CytC3 then catalyzes tandem chlorinations of the γ-methyl of Aba. Lastly, CytC4 is the thioesterase that hydrolytically releases chlorinated aminobutyrate from the phosphopantetheine arm of CytC2.

CytC3 is part of a newly discovered class of bacterial halogenase enzymes that use α-ketoglutarate (αKG) and non-heme Fe(II) for catalysis.[6] The only non-heme iron halogenase that has been structurally characterized thus far is the syringomycin biosynthetic enzyme II (SyrB2), which is involved in the biosynthesis of the anti-fungal agent syringomycin.[7-9] CytC3 is highly

similar to SyrB2, with 58% identity and 71% homology.[10] All αKG/Fe(II)-dependent enzymes share a common fold, referred to as a cupin fold, comprising anti-parallel β-strands in a jelly-roll motif.[11] The iron cofactor is located in the center of the jelly-roll and is typically coordinated by a 2-His-1-carboxylate (from Asp or Glu) triad.[12] Reactions catalyzed by this enzyme superfamily include epimerization, cyclization, epoxidation, and hydroxylation among others, where the hydroxylation reaction is the most common.[13] A major difference between hydroxylases and halogenases is that the hydroxylases coordinate the iron with the typical 2-His-1-carboxylate motif, whereas the halogenases coordinate the iron with only two histidines. Alanine is in place of the aspartate/glutamate of the 2-His-1-carboxylate motif in the halogenases, the small size of the alanine substitution and the non-coordinating nature of the side chain allow room for chloride to bind directly to iron in the position that is typically occupied by the side chain carboxylate.[8] In addition to the protein ligands, the remaining coordination sites of iron are taken by chloride, αKG and molecular oxygen.[13] Upon decarboxylation of αKG, an Fe(IV)-oxo species is generated, which is proposed to be the active intermediate responsible for hydrogen abstraction in both hydroxylases and halogenases.[14-17] The step after hydrogen abstraction is where the two mechanisms diverge; in hydroxylases, it is hypothesized that Fe(III)-OH undergoes a hydroxyl radical rebound to form the hydroxylated product.[14] In halogenases, the substrate radical is postulated to abstract a chlorine radical to form the chlorinated product.[17] It is as yet unclear how the proteins selectively promote hydroxyl or chloride transfer to the substrate.

Efforts have been made to convert the halogenase to hydroxylase and vice versa. Substitution of alanine with aspartate or glutamate in halogenase SyrB2, and substitution of aspartate for alanine in hydroxylase TauD abolishes all activity without the gain of new function in both cases.[8, 18] However, it is unclear if there is chloride binding to the active site iron in the

hydroxylase mutant. From a protein engineering perspective, it is important to understand the features responsible for chloride binding in these halogenases. However, with only one structurally characterized non-heme iron halogenase, SyrB2, it is difficult to pinpoint the essential attributes for chloride binding. In this study, the structure of a second halogenase, CytC3, was solved showing an open conformation of the active site. The comparison of the open conformation of CytC3 with the closed conformation of SyrB2 provides some insights into the features important for chloride binding.

IV.C. Materials and Methods

Overexpression, Purification and In Vitro Reconstitution of CytC3

The halogenase CytC3 was overproduced and purified as described[6] with the exception that the iron-reconstitution step was performed after the gel filtration. Following reconstitution, the protein was desalted on a Bio-Gel P-6 DG (Bio-Rad) column equilibrated in 20 mM Hepes, pH 7.5, 80 mM NaCl. Fractions containing proteins were pooled, concentrated in Amicon stirred ultrafiltration cells (Millipore) under N_2 pressure, flash-frozen in liquid N_2 in 25 μL aliquots, and stored at -80 °C.

Crystallization, Data Collection and Processing

CytC3 crystals were obtained using the hanging drop method by mixing 1 μL of protein (38 mg/mL CytC3 in 20 mM HEPES, pH 7.5) with 1 μL of 2.8 M sodium acetate trihydrate, pH 7.0 precipitant solution (Hampton Research), and placing the drop over 0.5 mL of precipitant solution at room temperature. αKG (Sigma) was added to the protein stock solution to a final concentration of 5 mM prior to crystallization setup. Square bi-pyramidal crystals appeared after

174

2 – 3 weeks. For crystals containing iron, the crystals grown aerobically without iron were brought into an anaerobic glove box (Coy Laboratory) and soaked in degassed mother liquor with $Fe(NH_4)_2(SO_4)_2$ (10 mM), αKG (5 mM) and chloride (80 mM). Variations on the order addition for $Fe(NH_4)_2(SO_4)_2$ and αKG cofactors include (1) soaked $Fe(NH_4)_2(SO_4)_2$ first, followed by αKG; (2) soaked αKG first, followed by $Fe(NH_4)_2(SO_4)_2$; and (3) soaked $Fe(NH_4)_2(SO_4)_2$ and αKG together. Prior to cryo-cooling in liquid nitrogen, crystals were soaked in mother liquor with 15% trehalose overnight. $Fe(NH_4)_2(SO_4)_2$, αKG, and chloride were also present in the mother liquor with the cryo-protectant when crystals were soaked overnight.

Data for the native crystals were collected to 2.2 Å resolution for both the apo and iron structures at Stanford Synchrotron Radiation Laboratory (Palo Alto, California) on beamline 11-1, integrated in Denzo, and scaled in Scalepack.[19] To locate the iron sites, a dataset was collected at the iron peak wavelength using crystals prepared as described above. An iron anomalous dataset was collected to 2.75 Å resolution at the Argonne Advanced Photon Source (Argonne, Illinois) on beamline NE-CAT-24-ID-C, integrated in Denzo, and scaled in Scalepack.[19] Data collection statistics are shown in Table IV.1.

Structure Determination and Refinement

The structure of apo CytC3 was solved by molecular replacement using PHASER[20] to 3 Å resolution. The structure of SyrB2 (2FCU), with 58% identity and 71% homology to CytC3, without any ligands or waters was used as a search model.[8] The best rotational and translational solution has correlation coefficient of 16 with two CytC3 molecules per asymmetric unit. The resulting electron density map was solvent flattened in SHARP.[21] Side chains were added to the model at 3 Å resolution followed by iterative rounds of model building and refinement using

XtalView and CNS, respectively.[22, 23] Phases were extended to 2.2 Å resolution, and waters were added to the model after few rounds of refinement, and further model building and refinements were done in COOT and CNS, respectively.[23, 24] No non-crystallographic symmetry (NCS) or sigma cutoff was used during the refinement.

Since the iron CytC3 crystals were isomorphous with the apo CytC3 crystals, a rigid body refinement followed by simulated annealing refinement was used. Further model building and refinements were done in COOT and CNS, respectively.[23, 24] No non-crystallographic symmetry (NCS) or sigma cutoff was used during the refinement. The iron structure was refined using a patch that restrains the bond lengths and angles of the ligands to the metal using higher weights for protein ligands (500 for both bond length and angle weights) and αKG (500 bond length weight and 100 angle weight), and lower weights for the water molecules (50 bond length weight and no angle weight).

Composite omit maps calculated using CNS[23] simulated annealing were used to validate the models, and an anomalous difference map calculated using CCP4[25] program was used to verify the positions of the irons. Refinement statistics for both models are shown in Table IV.2.

Final Protein Models

The final apo structure was refined to 2.2 Å resolution with R_{cryst} 21.8% and R_{free} 26.0%. The final iron structure was refined to 2.2 Å resolution with R_{cryst} 23.8% and R_{free} 27.8% with the same reflection test set as for the apo structure. For both the apo and iron structures, residues 8 – 317 were observed out of a total of 319 residues. There is one major chain break in both structures (residues 178 – 220 in molecule A, and 179 – 216 in molecule B), and a minor chain break only in molecule A of the apo structure (residues 170 – 172). In the apo structure, 89.2%

of the residues are in the most-favored region of the Ramachandran plot, with 0.2% of the residues in the disallowed region. In the iron structure, 86.4% of the residues are in the most-favored region of the Ramachandran plot, with 0.9% of the residues in the disallowed region.

IV.D. Results

Structural Overview

We have determined the structures of apo and iron/αKG-bound CytC3, both at 2.2 Å resolution. Both structures belong to $P4_32_12$ space group, and the asymmetric unit contains a dimer of two CytC3 molecules related by a two-fold non-crystallographic symmetry. Although CytC3 crystallizes as a dimer, gel filtration chromatography indicates that CytC3 functions as a monomer.[6] The core structure of CytC3 is a β-sandwich 'jelly-roll' motif, common to the αKG non-heme iron enzyme family, and it comprises beta-strands β2, β3, β5, β6, and β9-12 (Figure IV.2.a). Dimerization occurs between beta-strand 3 of each monomer, each strand hydrogen bonding to form a continuous eight-stranded beta-sheet (consists of β3, β6, β9 and β11 from each monomer). In addition to the β-strands, there are five alpha helices near the N-terminus packed on the outside of the dimer (Figure IV.2.a). The apo and iron/αKG-bound CytC3 structures are very similar, with an RMSD of 0.461 Å over 263 C_α atoms.

Each monomer in the CytC3 dimer has a large disordered region encompassing residues 178 to 219, which we will refer to as the "missing loop" (Figure IV.2.a). There is electron density around the vicinity of the missing loop indicating the presence of mobile amino acids; however, the quality of the density is too poor to model. Residues near the ends of the missing loop have elevated B-factors comparing to the rest of the protein, 75 Å^2 compared to 45 Å^2. Structural alignment between CytC3 and SyrB2, with an RMSD of 1.16 Å over 221 C_α atoms,

predicts that two beta-strands ($\beta7$ and $\beta8$) should be located in the missing loop (Figure IV.2.b and 3). These two beta-strands and the residues of the corresponding missing loop are very important in SyrB2 as they cover the active site.[8]

The structural alignment between CytC3 and SyrB2 strongly indicates that the interactions between the two molecules of CytC3 in the asymmetric unit are a crystallographic artifact. The core secondary structural elements in both CytC3 and SyrB2 align well, with the exception of the two β-strands and the two 3_{10} helices that cover the active site in SyrB2 that are missing in CytC3. This part of the structure should occupy the space that is occupied by the second molecule in the CytC3 non-crystallographic dimer (Figure IV.2.b). Therefore, crystal packing has allowed us to capture an open active site conformation of the CytC3 protein. Based on sequence alignment, it is predicted that in solution the inter-molecular hydrogen bonds across the two CytC3 molecules are not present, and the missing loop in CytC3 folds over the active site as it does in SyrB2 (Figure IV.3). Given that dimerization of CytC3 requires shifting the missing loop (residues 178 – 220) out of the way, one can imagine that this region of the structure has some mobility and it may be important in interacting with the substrate, which is the amino-acylated thiolation domain (CytC2).

Active Site Features

The active sites of the two CytC3 monomers are located at the dimer interface positioning them very close to each other; the irons are within 10 Å. Surface representations reveal that the active sites are highly solvent exposed from one side of the dimer (Figure IV.2.c). Unlike most other enzymes in the cupin superfamily, CytC3 and SyrB2 only contain two histidine residues as protein ligands to the iron.[8] Other iron coordination sites are available to bind chloride, waters

and/or αKG. Since co-crystallization with iron, chloride and αKG yielded slow growing apo crystals, we obtained a structure with iron and αKG bound by soaking the apo CytC3 crystals in high concentrations of $Fe(NH_4)_2(SO_4)_2$ (10 mM), αKG (5 mM) and chloride (80 mM). In the soaked structure, both CytC3 molecules have one iron bound; however, even with such high concentrations of external ligands, only one molecule in the dimer has αKG in the active site, and no chloride is observed in either molecule.

In addition to the iron and the iron-bound αKG, the open and solvent accessible active sites of CytC3 can accommodate other molecules from the crystallization conditions such as excess αKGs, acetate and sulfate. There are many charged residues near the active site that interact with these molecules. For example, the side chains of R102 and D116 and the backbone of H125 and F123 interact with an extra molecule of αKG, and a sulfate ion hydrogen bonds with D116 and K62 (Figure IV.4.a). It is unlikely that these excess binding events are physiologically relevant.

To verify the presence of iron in the active site, a data set was collected at the iron peak wavelength, 1.73989 Å. One iron per monomer is observed at the active site as judge by the 12σ peak in the anomalous difference map (Figure IV.4.b). Additionally, if water molecules are placed in the iron sites and refined after fixing the B-factor to the average B-factors of the protein at 45 $Å^2$, the resultant occupancies for waters are 3.19 and 2.91, suggesting the presence of a heavier atom at the site. The occupancies for iron atoms refined in the iron sites are 0.93 and 0.84, which are consistent with the conclusion based on the anomalous difference map.

Chloride is observed bound to iron in the SyrB2 structure, an interaction that is expected to be important for the chlorination activity of the protein.[8] Since CytC3 and SyrB2 are structurally similar to each other and perform the same chemistry, it is predicted that the

mechanism by which CytC3 chlorinates is the same as that for SyrB2. Therefore, a chloride ion is expected to be one of the ligands for the iron. To examine whether water or chloride is bound to iron, B-factors were refined for modeled water and modeled chloride after fixing occupancies to one. The B-factors for chloride ions modeled as ligands are 185.3 Å^2 and 178.3 Å^2, which are much higher than the average B-factors for the protein, suggesting that an atom lighter than chloride is bound. The B-factors for modeled waters are 43.5 Å^2 and 32.2 Å^2, which are consistent with the average B-factors for the rest of the protein, indicating the presence of water molecules in the other open iron coordination sites (Figure IV.4.c). It is surprising that no chloride ion is bound to the iron given the concentration of chloride used in the soaks was 80 mM. In contrast, iron-containing SyrB2 was unable to be crystallized without a chloride bound.[8]

The composite omit map of the active site only shows electron density for αKG in one of the CytC3 monomers (Figure IV.4.c, data for the monomer without αKG is not shown), suggesting the open active site conformation of CytC3 is not an optimal environment for αKG binding. The B-factor for the active site αKG is also higher than the average B-factor for the protein, 67 Å^2 versus 45 Å^2, indicating that the αKG is not as well ordered. The αKG binds to the iron through its O2 and O5 oxygens (Figure IV.4.b for numbering nomenclature) in a bidentate and planar fashion. Bidentate binding of αKG to iron is strictly conserved in the αKG/Fe(II)-dependent enzyme family, although the geometry of αKG varies.[26]

CytC3 and SyrB2 Active Site Comparison

Our results suggest that residues that do not directly contact the iron are also important in binding αKG and chloride in these halogenating enzymes. These surrounding residues are highly conserved between CytC3 and SyrB2; however, due to the difference in crystal lattice contacts,

the interactions of these CytC3 residues with αKG are not the same as those observed in SyrB2. If the structures of CytC3 and SyrB2 are aligned by all the residues on one side of the active site (H125, R259, T148, F106, W150, R253, and K108 in CytC3, and N123, R254, T143, F104, W145, R248 and K106 in SyrB2), the agreement of the aligned residues is good. However, the agreement between the unaligned residues (H118, H240 and T115 for CytC3, and H116, H235 and T113 for SyrB2) is not (Figure IV.5.a). Likewise, when residues on the other side of the active site (T115 in CytC3, T113 in SyrB2) and (H118 and H240 in CytC3, H116 and H235 in SyrB2) are aligned, the upper half of the active site in CytC3 is no longer superimposed with that in SyrB2 (Figure IV.5.b). Collectively, these differences dramatically affect the contacts between protein, αKG and chloride.

There are a number of interactions that anchor the αKG in the SyrB2 active site, such as hydrophobic stacking against F104 and many hydrogen binding interactions (Figure IV.5.c). Hydrophobic stacking between αKG and the protein is not observed in CytC3 due to the slightly altered conformation of αKG, and the greater distance between the αKG and the corresponding phenylalanine (F106) (Figure IV.5.d). There are also multiple hydrogen bonds between surrounding residues and αKG in SyrB2: R254 interacts with O2 of αKG (Figure IV.4.b for αKG numbering nomenclature), W145 and R248 interact with O3 of αKG, R248 and T115 interact with O4 of αKG (Figure IV.5.c). However, none of these interactions are preserved in the "open" CytC3 active site; in each case, the distances are too long to represent hydrogen bonds (Figure IV.5.d). Instead, new interactions are observed in this conformation of the CytC3 active site to stabilize the αKG: K108 interacts with O4 of αKG, and W150 interacts with O1 of αKG through a water molecule (w3) (Figure IV.5.d). Overall, the "open" active site conformation of CytC3 results in αKG having fewer total stabilizing interactions compared to

SyrB2, which is consistent with the higher degree of disorder observed for this αKG molecule. We predict that these weak interactions between αKG and CytC3 are only present when the active site is in a relatively open state as observed in this structure, and that αKG binding will mimic that in SyrB2 in a closed structure of the active site as would occur during catalysis.

The comparison of CytC3 and SyrB2 active sites provides insight into why no chloride is bound to iron in CytC3. In SyrB2, a hydrogen-bonding network between residues N123, T143, R254 and chloride connected by water molecules appears to stabilize the chloride in the active site (Figure IV.5.c). However, no such hydrogen-bonding network is observed in CytC3, as the corresponding residues H125, T148 and R259 are too far from the chloride binding site for a single water molecule to form a hydrogen-bonding network. Additionally, the chloride in SyrB2 is situated in a large hydrophobic pocket, consisting of A118, F121 and the β-carbon of S231 (Figure IV.6.a), that might be important for its binding.[8] This hydrophobic pocket is not present in this open structure of CytC3; while the alanine is in the same relative place, both the corresponding phenylalanine (F123) and serine (S236) residues are more than 12 Å and 8 Å away, respectively (Figure IV.6.b).

There are three chloride-binding loops in SyrB2 (55 – 61, 122 – 128 and 271 – 275) that undergo conformational changes to close the active site upon chloride binding.[8] Based on the structural alignment between CytC3 and SyrB2, the predicted chloride-binding loops in CytC3 consists of residues 60 – 63, 124 – 130 and 278 – 280 (Figure IV.3). The first predicted chloride-binding loop in CytC3 is shortened by a three residue gap compared to that of SyrB2, and residue 63 is part of α-helix 3 instead of being part of a loop. The second predicted chloride-binding loop in CytC3 precedes β-strand 4 in the structure, as is the case for SyrB2. The third predicted chloride-binding loop in CytC3 is located at a hairpin between β-strands 14 and 15. No

movement in these predicted chloride-binding loops is observed when the apo and the Fe(II)/αKG/Cl⁻ soaked CytC3 structures are compared, which is not surprising since no chloride is bound to the iron center in CytC3. It is difficult to confirm the locations, the lengths or the roles of these chloride-binding loops in CytC3 without a chloride bound structure for comparison.

IV.E. Discussion

The metal centers in CytC3 and SyrB2 vary from the common iron centers found in other members of the αKG-mononuclear iron enzyme family. These halogenases harvest the oxidizing power of the simple metal center of other αKG/Fe(II)-dependent enzymes and adapt it to bind chloride for chlorination of unactivated carbons. The metal centers in hydroxylases with three-protein ligands (two His and one carboxylate from Asp or Glu) are transformed to a center with two-protein ligands (only two His) in halogenases; in the place of the carboxylate, an alanine residue is present in the halogenases, leaving room for chloride to bind to the iron.[27] There has been much interest in understanding the features that are important for chloride binding to an iron center, converting hydroxylation chemistry to that of halogenation. Is it a simple steric issue or are there other factors involved, such as hydrogen-bonding or hydrophobic interactions through the secondary coordination sphere of the metal? The structural comparison of a halogenase containing no bound chloride with one containing bound chloride has allowed us to investigate these questions. Although the CytC3 used in these studies is fully active, a crystalline state of this enzyme has been captured to which chloride binding is impaired. While in solution this "open" chloride-free state would be in equilibrium with a "closed" chloride-bound state, in

the crystal, the transformation between states is prevented by lattice contacts allowing us to observe this otherwise transient enzyme conformation.

With only two protein ligands in iron halogenases, iron does not bind tightly in the absence of αKG. When apo CytC3 crystals are soaked in 10 mM $Fe(NH_4)_2(SO_4)_2$, or soaked in 10 mM $Fe(NH_4)_2(SO_4)_2$ first, followed by 5 mM αKG, or vice versa, little density is observed in the active site for αKG or iron, indicating very low occupancies of both. Iron is only bound in high occupancy to the crystalline CytC3 active site when the crystals are soaked in both 10 mM $Fe(NH_4)_2(SO_4)_2$ and 5 mM αKG together, which suggests the binding of αKG helps to keep iron in the active site. Poor iron binding is also observed in the D101A mutant of TauD, where the iron content is low (0.50 ± 0.49 mole Fe/mole subunit) when it is measured without the presence of αKG.[18] Like CytC3, SyrB2 can only be reconstituted with 0.3 – 0.4 mole Fe/mole subunit in the absence of αKG, but the iron content is much higher (0.85 – 0.95 mole Fe/mole subunit) in the presence of αKG.[7]

The hydrogen-bonding network between residues surrounding the active site and the chloride appears to be an important feature for chloride binding in CytC3 and SyrB2. The "closed" active site in SyrB2 allows residues near the metal center (R254, T143 and N123) to hydrogen bond with chloride through two water molecules. In the light-driven chloride pump halorhodopsin, the chloride is also stabilized in a hydrogen-bonding network with protein residues via water molecules.[28] However, due to the "open" active site in this CytC3 structure, the corresponding residues are too far from the chloride binding site for hydrogen bonding, which may contribute to the lack of observed chloride binding. Secondary coordination sphere hydrogen bonds have been shown to be important for modulating ligand affinity to metal centers in both metalloproteins as well as synthetic model compounds that mimic enzymatic active sites.

For example, the removal of a hydrogen bond distal to the active site in horseradish peroxidase changes the metal center from five coordinate to six coordinate in the resting state.[29] Likewise, in a series of model compounds with varying numbers of secondary coordination sphere hydrogen bonds, a range of affinities for oxygen binding to cobalt is observed.[30]

Another feature that may be important for chloride binding to iron is the presence of a suitable hydrophobic pocket around the chloride. In the closed SyrB2 active site, hydrophobic residues (Ala, Phe and the β-carbon of Ser) are within 4 Å of chloride, providing a hydrophobic pocket for chloride binding. Corresponding residues in the open CytC3 active site are more than 8 Å to 12 Å from the chloride, and are thereby unable to provide a similar environment. Hydrophobic chloride binding sites have been observed in the light driven chloride pump halorhodopsin and ClC chloride binding channel as well. In both of these proteins, the chloride binding environment is made up of hydrophobic residues such as Gly, Ile, Phe, Trp, the β-carbon of Ser, and the γ-carbon of Thr.[28, 31] Along the lines of a binding pocket for chloride, Tolman, Que and co-workers have recently explored the use of a sterically bulky α-ketocarboxylate ligands to prepare structural models of iron-halogenase active sites.[32] They found that the use of bulky α-ketocarboxylate ligands leads to inactive iron complexes that structurally mimic halogenase active sites, complete with an Fe-Cl bond, but are unable to catalyze oxidative decarboxylation of α-keto ligand. On the other hand, the use of less bulky α-ketocarbxylate ligands leads to the formation of iron complexes that lack an Fe-Cl bond, but are able to catalyze oxidative decarboxylation of the α-keto ligand. Thus, while steric hinderance is useful for the creation or stability of the Fe-Cl bond in these models, it impedes decarboxylation activity, presumably by preventing the attack of the bound O_2 on the α-keto group. The challenge will be

to create the appropriate binding pocket for chloride without interfering with the oxidative decarboxylation reaction.

Since the discovery of αKG/Fe(II) dependent halogenases, several groups have tried to convert hydroxylases to halogenases and vice versa. In an attempt to convert SyrB2 to do hydroxylation chemistry, A118D and A118E mutant proteins were prepared. These mutations abolished chlorination activity, but no hydroxylation activity was observed.[8] Hausinger *et al.* tried to convert hydroxylase TauD to do a chlorination reaction by mutating the active site carboxylate to an alanine, but the D101A variant showed no chlorination activity.[18] However, in the latter case, it is unclear if the lack of chlorination was due to a lack of chloride and/or iron binding to the active site or if the mutation introduced a perturbation to the protein structure. The data presented here suggest that the selective activity towards hydroxylation or chlorination depends on more than just the steric issue of the number of protein ligands to iron. Based on our observations in the open CytC3 active site structure, chloride binding to iron depends on a combination of a hydrogen-bonding network and/or hydrophobic environment, in addition to an open coordination at the iron site.

IV.F. Conclusion

The fortuitous "open" active site structure of CytC3 has allowed us to investigate the protein features that allow chloride to bind directly to the iron center, a necessary step for chlorination and not hydroxylation in this family of enzymes. The structure of SyrB2 revealed that the substitution of Ala for Asp/Glu creates space in the protein structure for chloride to bind directly to iron, however, as the biochemistry shows, this substitution is insufficient to switch a hydroxylase to a halogenase. Therefore, the SyrB2 structure alone did not indicate which other

protein features might be important for chloride binding. Here we find that chloride binding is impaired when a hydrogen-bonding network to the surrounding residues is interrupted, and when a hydrophobic pocket is not properly formed. These findings will assist in the protein design of halogenating enzymes and in the creation of model complexes.

IV.G. Future Work

Although the structure of CytC3 provided insight into the necessary scaffold of a halogenase active site, it provides no information about the interactions between the substrate and the enzyme active site. Information about these interactions could explain how this class of enzymes achieves regio- and stereo-specific halogenation, which is essential for engineering new therapeutics. Therefore, the future extension of this study will be to crystallize CytC3 with its substrate holo-CytC2 and amino acylated substrate analogs. Screening for the CytC3/CytC2 complex has been initiated.

CytC2 and CytC3 Purification and CytC2/CytC3 Complex Formation

The apo form of the thiolation domain protein CytC2 was overproduced and purified as described[6] with the following modification to the protocol. Following elution from the Ni-NTA resin, proteins were loaded onto a 26/60 Superdex 75 gel filtration column (Amersham Biosciences), the apo CytC2 was eluted in 20 mM HEPES pH 7.5, concentrated, frozen in liquid nitrogen and stored at -80 °C. The holo form of CytC2 was prepared via incubation of 335 μM apo CytC2 with 15 mM $MgCl_2$, 1.5 mM coenzyme A and 3.7 μM phosphopantetheinyl transferase Sfp in 20 mM HEPES, pH 7.5, at 23 °C for 1.5 h. The mixture was subsequently

concentrated and then loaded onto a 26/60 Superdex 75 column equilibrated in 20 mM HEPES, pH 7.5. The eluted holo-CytC2 was concentrated, frozen in liquid nitrogen and stored at -80 °C.

CytC3 purification is described in the Materials and Methods section. Four valine amino acylated CytC2 analogs were also synthesized by our collaborator Danica G. Fujimori in Professor Christopher Walsh's laboratory (Figure IV.7).

Prior to preparing the CytC2/CytC3 complex, CytC3 was reconstituted with Fe(NH$_4$)$_2$(SO$_4$)$_2$ (2 mM) and αKG (5 mM) in storage buffer (10 mM HEPES, pH 7.5, and 80 mM NaCl storage buffer). Reconstituted CytC3 and holo-CytC2 proteins were mixed at a 1:2 molar ratio to form three mixtures of CytC2/CytC3 complexes at three different final concentrations: (1) 2.3 mg/mL CytC2 with 3.5 mg/mL CytC3, (2) 6.5 mg/mL CytC2 with 10 mg/mL CytC3, and (3) 11 mg/mL CytC2 with 17 mg/mL CytC3.

CytC2/CytC3 Complex Crystal Screening

With three different concentrations of the CytC2/CytC3 complex, sparse matrix screens were carried out with 1:1 drop to precipitant ratio using hanging drop methods. The following sparse matrix screens were used: Index, Crystal Screen I and II, Wizard Screen I, II and III, PEG/ION, Salt screen, ProComplex Screen, and Pre-Synergy Screen. These screens were set up both at room temperature in an anaerobic chamber (Coy Scientific), as well as at 4°C in an anaerobic chamber (Coy Scientific).

Crystallization screens with the CytC2/CytC3 complex mixture 1 were clear, and a few of the drops showing oil droplets or phase separation. Crystallization screens with the CytC2/CytC3 complex mixture 3 were mostly heavy precipitate, suggesting that the protein concentration was too high. Crystallization screens with the CytC2/CytC3 complex mixture 2 yielded three hits

summarized in Table IV.3. Index #4 was only observed once, only with the CytC2/CytC3 complex mixture 1. Needle cluster crystals appeared after 2-3 days, but disappeared after 5 days. This hit was never reproduced. ProComplex #27 and #34 yielded the same type of crystal at 4°C with CytC2/CytC3 complex mixture 2, but neither of these hits were reproducible. Index #66 and #67 hits were observed both at room temperature and at 4°C using CytC2/CytC3 complex mixture 2. At room temperature, both the Index #66 and #67 condition yielded needle cluster crystals grown from precipitate after 2-3 days (Figure IV.8). At 4°C, the same needle cluster crystals appeared in clear drops after 1-2 weeks in Index #66, and the same crystals appeared in slightly precipitated drops after 1-2 weeks in Index #67. Index #66 was optimized at 4°C by varying buffers pH (pH 5-7, with 0.2 pH increments), ammonium sulfate concentration (0.15-0.3 M, with 0.05M increments) and PEG 3,350 concentrations (20%-35%, with 2% increments). In addition, different salts, PEGs and buffers were used to replace ammonium sulfate and PEG 3,350 in an attempt to improve crystals (Table IV.4). Drop size ratios and sitting drops were also tested in an attempt to improve the crystal quality; however neither the drop size ratio nor using the sitting drop method improved the crystals.

The best optimized CytC3/CytC2 complex crystals were obtained using the hanging drop method at 4°C in an anaerobic chamber by mixing 1 µL of CytC2/CytC3 complex mixture 2 with 1 µL of 0.1 M Bis-Tris, pH 5.5, 0.2 M ammonium sulfate, and 30% PEG 3,350, and placing the drop over 0.5 mL of precipitant solution. Both holo-CytC2 and four amino acylated CytC2 analogs were used in the optimization process, and the crystals yielded by either holo-CytC2 or any of the amino acylated CytC2 analogs were the same. Needle haystack crystals appeared from clear drops after 1-2 weeks like the one shown in Figure IV.8.a. In order to confirm that these crystals are crystals of the CytC3/CytC2 complex, a few clusters of these needle haystacks were

washed in mother liquor three times to remove any protein on the surface of the crystal, and dissolved in water to be analyzed in SDS-PAGE electrophoresis. The crystals grown under this condition were confirmed to be crystals of CytC3/CytC2 complex (Figure IV.8.b). After several rounds of streak seeding, needle haystack crystals were turned into thin plate crystals. Crystals were cryo-cooled in liquid nitrogen without using any cryo-protectant.

Data Collection and Processing

A preliminary dataset of the thin plate CytC3/CytC2 crystal was collected to 3.5-4 Å resolution at the Advanced Photon Source at the Argonne National Laboratory (Argonne, Illinois) on beamline NE-CAT-24-ID-C. Using one frame, the crystal was indexed using HKL2000 to a primitive orthorhombic space group with cell dimension 48.8, 73.5, 106.6 Å.[19] However, the crystal probably has multiple lattices, as judged by the split spots on the diffraction pattern (Figure IV.9), and was not processed. Further data processing and/or obtaining of better crystals will be necessary to yield a structure of the complex of two NRPS domains.

IV.H. Acknowledgements

IV.I. References

1. Walsh, C., Molecular mechanisms that confer antibacterial drug resistance. *Nature* **2000**, 406, (6797), 775-81.

2. Fenical, W.; Jensen, P. R., Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat Chem Biol* **2006**, 2, (12), 666-73.

3. Bender, C. L.; Alarcon-Chaidez, F.; Gross, D. C., Pseudomonas syringae phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **1999**, 63, (2), 266-92.

4. Grgurina, I.; Barca, A.; Cervigni, S.; Gallo, M.; Scaloni, A.; Pucci, P., Relevance of chlorine-substituent for the antifungal activity of syringomycin and syringotoxin, metabolites of the phytopathogenic bacterium Pseudomonas syringae pv. syringae. *Experientia* **1994**, 50, (2), 130-3.

5. Schnarr, N. A.; Khosla, C., Biological chemistry: just add chlorine. *Nature* **2005**, 436, (7054), 1094-5.

6. Ueki, M.; Galonic, D. P.; Vaillancourt, F. H.; Garneau-Tsodikova, S.; Yeh, E.; Vosburg, D. A.; Schroeder, F. C.; Osada, H.; Walsh, C. T., Enzymatic generation of the antimetabolite gamma,gamma-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chem Biol* **2006**, 13, (11), 1183-91.

7. Vaillancourt, F. H.; Yin, J.; Walsh, C. T., SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc Natl Acad Sci U S A* **2005**, 102, (29), 10111-6.

8. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006**, 440, (7082), 368-71.

9. Vaillancourt, F. H.; Vosburg, D. A.; Walsh, C. T., Dichlorination and bromination of a threonyl-S-carrier protein by the non-heme Fe(II) halogenase SyrB2. *Chembiochem* **2006**, 7, (5), 748-52.

10. Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.; Thompson, J. D., Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **2003**, 31, (13), 3497-3500.

11. Dunwell, J. M.; Purvis, A.; Khuri, S., Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **2004**, 65, (1), 7-17.

12. Hegg, E. L.; Que, L., Jr., The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* **1997**, 250, (3), 625-9.

13. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J., Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *Journal of Inorganic Biochemistry* **2006,** 100, (4), 644-669.

14. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M.; Krebs, C., The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: A high-spin Fe(IV) complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli (vol 42, pg 7497, 2003). *Biochemistry* **2004,** 43, (4), 1134-1134.

15. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine :alpha-ketoglutarate dioxygenase (TauD). *Journal of the American Chemical Society* **2003,** 125, (43), 13008-13009.

16. Bollinger, J. M.; Price, J. C.; Barr, E. W.; Tirupati, B.; Krebs, C., Characterization of a high-spin Fe(IV) intermediate in the reaction of taurine/alpha-ketoglutarate dioxygenase (TauD). *Journal of Inorganic Biochemistry* **2003,** 96, (1), 63-63.

17. Galonic, D. P.; Barr, E. W.; Walsh, C. T.; Bollinger, J. M., Jr.; Krebs, C., Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat Chem Biol* **2007,** 3, (2), 113-6.

18. Grzyska, P. K.; Muller, T. A.; Campbell, M. G.; Hausinger, R. P., Metal ligand substitution and evidence for quinone formation in taurine/alpha-ketoglutarate dioxygenase. *J Inorg Biochem* **2007,** 101, (5), 797-808.

19. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **1997,** 276, 307-326.

20. Read, R. J., Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D-Biological Crystallography* **2001,** 57, 1373-1382.

21. delaFortelle, E.; Bricogne, G., Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Macromolecular Crystallography, Pt A* **1997,** 276, 472-494.

22. McRee, D. E., XtalView Xfit - A versatile program for manipulating atomic coordinates and electron density. *Journal of Structural Biology* **1999,** 125, (2-3), 156-165.

23. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **1998,** 54, 905-921.

24. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **2004,** 60, 2126-2132.

25. Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **2003,** 59, (Pt 7), 1131-7.

26. Hausinger, R. P., FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **2004,** 39, (1), 21-68.

27. Straganz, G. D.; Nidetzky, B., Variations of the 2-His-1-carboxylate theme in mononudear non-heme Fe-III oxygenases. *Chembiochem* **2006,** 7, (10), 1536-1548.

28. Kolbe, M.; Besir, H.; Essen, L. O.; Oesterhelt, D., Structure of the light-driven chloride pump halorhodopsin at 1.8 angstrom resolution. *Science* **2000,** 288, (5470), 1390-1396.

29. Mukai, M.; Nagano, S.; Tanaka, M.; Ishimori, K.; Morishima, I.; Ogura, T.; Watanabe, Y.; Kitagawa, T., Effects of concerted hydrogen bonding of distal histidine on active site structures of horseradish peroxidase. Resonance Raman studies with Asn70 mutants. *Journal of the American Chemical Society* **1997,** 119, (7), 1758-1766.

30. Lucas, R. L.; Zart, M. K.; Murkerjee, J.; Sorrell, T. N.; Powell, D. R.; Borovik, A. S., A modular approach toward regulating the secondary coordination sphere of metal ions: Differential dioxygen activation assisted by intramolecular hydrogen bonds. *Journal of the American Chemical Society* **2006,** 128, (48), 15476-15489.

31. Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R., X-ray structure of a CIC chloride channel at 3.0 angstrom reveals the molecular basis of anion selectivity. *Nature* **2002,** 415, (6869), 287-294.

32. Friese, S. J.; Kucera, B. E.; Young, V. G., Jr.; Que, L., Jr.; Tolman, W. B., Iron(II) complexes of sterically bulky alpha-ketocarboxylates. Structural models for alpha-ketoacid-dependent nonheme iron halogenases. *Inorg Chem* **2008,** 47, (4), 1324-31.

IV.J. Table and Figures

Table IV.1. Data collection and processing statistics for apo, iron and iron anomalous CytC3 datasets.

Structure	Apo CytC3	Iron CytC3	Iron Anomalous CytC3[c]
Crystal Parameters			
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit cell dimension (Å)	90.0, 90.0, 249.2	89.0, 89.0, 248.4	89.8, 89.8, 249.1
# of molecules per ASU	2	2	2
Data Collection			
Wavelength (Å)	1.0	1.0	1.73989
Resolution limit[b] (Å)	50 – 2.2 (2.28 – 2.20)	50 – 2.2 (2.28 – 2.20)	50 – 2.75 (2.85 – 2.75)
R_{sym}[a,b] (%)	5.6 (20.0)	8.2 (48.0)	9.8 (35.9)
No. of unique reflections	50903	50583	50102
Redundancy[b]	13.7 (2.6)	10.9 (7.1)	11.0 (2.1)
Completeness[b] (%)	99.1 (81.5)	97.6 (81.3)	98.3 (91.2)
I/sigma[b]	33.8 (3.7)	27.8 (2.5)	11.3 (2.3)

[a]$R_{\text{sym}} = (\Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|)/ \Sigma_{hkl}\Sigma_iI_i(hkl)$ for n independent reflections and i observations of a given reflection. $<I(hkl)>$ = average intensity of the ith observation.

[b]Numbers for the highest resolution shell are shown in parentheses.

[c]An iron anomalous CytC3 dataset was collected at iron peak wavelength and scaled anomalously.

194

Table IV.2. Model refinement statistics for apo and iron CytC3 structures.

Structure	Apo CytC3	Iron CytC3
$R_{\text{cryst}}/R_{\text{free}}$[a] (%)	21.8/26.0	23.8/27.8
No. of protein atoms	4313	4321
No. of metal ion	0	2
No. of ligands	5	13
No. of water molecules	249	247
Disordered residues	170-172, 178-220 (chain A) 179-215 (chain B)	178-220 (chain A) 179-216 (chain B)
Rmsd bond lengths (Å)	0.0065	0.0077
Rmsd bond angles (°)	1.33	1.43
Ramachandran analysis		
Most favored (%)	89.2	86.4
Allowed (%)	9.7	12.1
Generously allowed (%)	0.9	0.7
Disallowed (%)	0.2	0.9
B-factors (Å^2)		
Protein	41	46
Iron	-	49
αKG	-	67
Water	45	50

[a]$R_{\text{cryst}} = \Sigma_h||F_o(h)| - |F_c(h)|| / \Sigma_h|F_o(h)|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is calculated the same way with a test set of reflections (10%) that are not used during refinement.

Table IV.3. Summary of initial crystal hits for CytC2/CytC3 complex.

Screen & Hit #	Condition	CytC2/CytC3 complex mixture	Temperature
Index #4	0.1 M Bis Tris, pH 6.5 2 M $(NH_4)_2SO_4$	1	Room temperature
Index #66	0.1 M Bis Tris, pH 5.5 0.2 M $(NH_4)_2SO_4$ 25% PEG 3,350	2	Room temperature, 4°C
Index #67	0.3 M Bis Tris, pH 6.5 0.4 M $(NH_4)_2SO_4$ 25% PEG 3,350	2	Room temperature, 4°C
ProComplex (Qiagen) #27	0.1 M Na acetate, pH 5 0.2 M ammonium acetate 20% PEG 4,000	2	4°C
ProComplex (Qiagen)	0.1 M MES, pH 5.5 0.15 M $(NH_4)_2SO_4$ 25% PEG 4,000	2	4°C

Table IV.4. A list of salts, PEGs and buffers in optimizing crystals of CytC2/CytC3 complex. These salts, PEGs and buffers were used to substitute for ammonium sulfate, PEG 3,350 and Bis Tris pH 5.5 in the original hit (Index #66) in an attempt to improve crystal quality. Each salt was mixed with all three PEGs and all four buffers.

Salts	PEGs	Buffers
0.2 M ammonium sulfate	25% PEG 400	0.1 M Bis Tris, pH 5.5
0.2 M sodium acetate	25% PEG 1,000	0.1 M Bis Tris, pH 6.5
0.2 M sodium chloride	25% PEG 3,350	0.1 M cacodylate, pH 6.5
0.4 M diammonium citrate		0.1 M imidazole, pH 6.5
0.2 M triammonium citrate		
0.2 M trisodium citrate		
0.2 M magnesium formate		
0.2 M ammonium nitrate		
0.2 M sodium nitrate		
0.2 M tacimate		

Figure IV.1. Production of 4-Cl-L-Aba by CytC1-4. CytC1 loads L-Aba-AMP onto the phosphopantetheine arm (wavy line) on CytC2. CytC3 chlorinates the tethered substrate to form 4-Cl-L-Aba. CytC1 is an adenylation domain (A) and CytC2 is a thiolation domain (T).
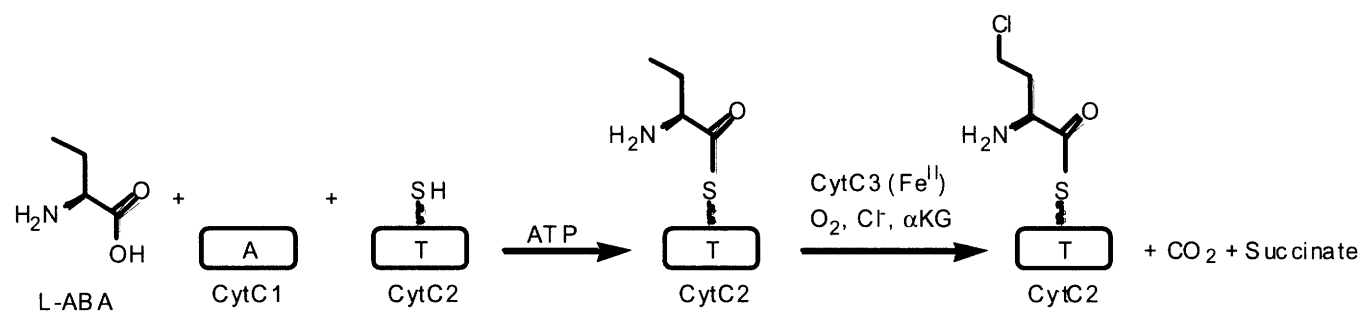
Figure IV.2. Overall structure of CytC3 dimer. a) Crystallographic dimer is colored by molecule. The iron ligands: two protein residues (His118 and His240) and αKG are shown in stick representation, and waters in active site are shown in sphere representation. The ends of the disordered region are indicated as "missing loop", which consists of residues 178-219. b) Structural alignment of SyrB2 monomer (pink) and CytC3 dimer (blue). c) Surface representation of CytC3 dimer showing access to active site on one face of the crystallographic dimer. Active site waters are colored in red; each monomer of CytC3 is colored yellow and blue. The ends of the disordered region are colored in green.
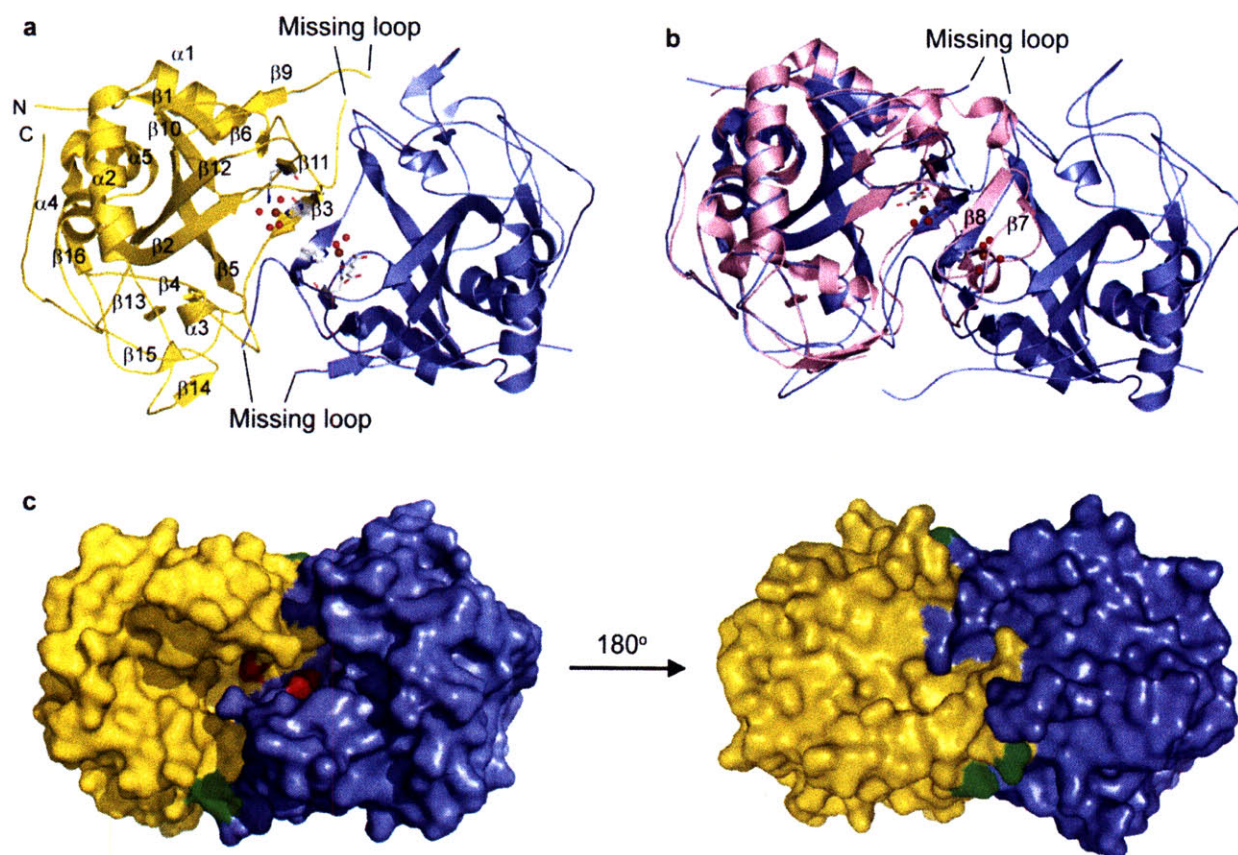
Figure IV.3. Sequence alignment of CytC3 and SyrB2 with secondary structures defined above the sequence. Alpha helices are indicated by helical lines with α labels, beta strands are indicated by arrows with β labels, and 3$_{10}$ helices are indicated by helical lines with η labels. Red stars indicate the active site residues, blue arrows indicate the ends of the missing loop, and green bars indicate the chloride binding loops in SyrB2.
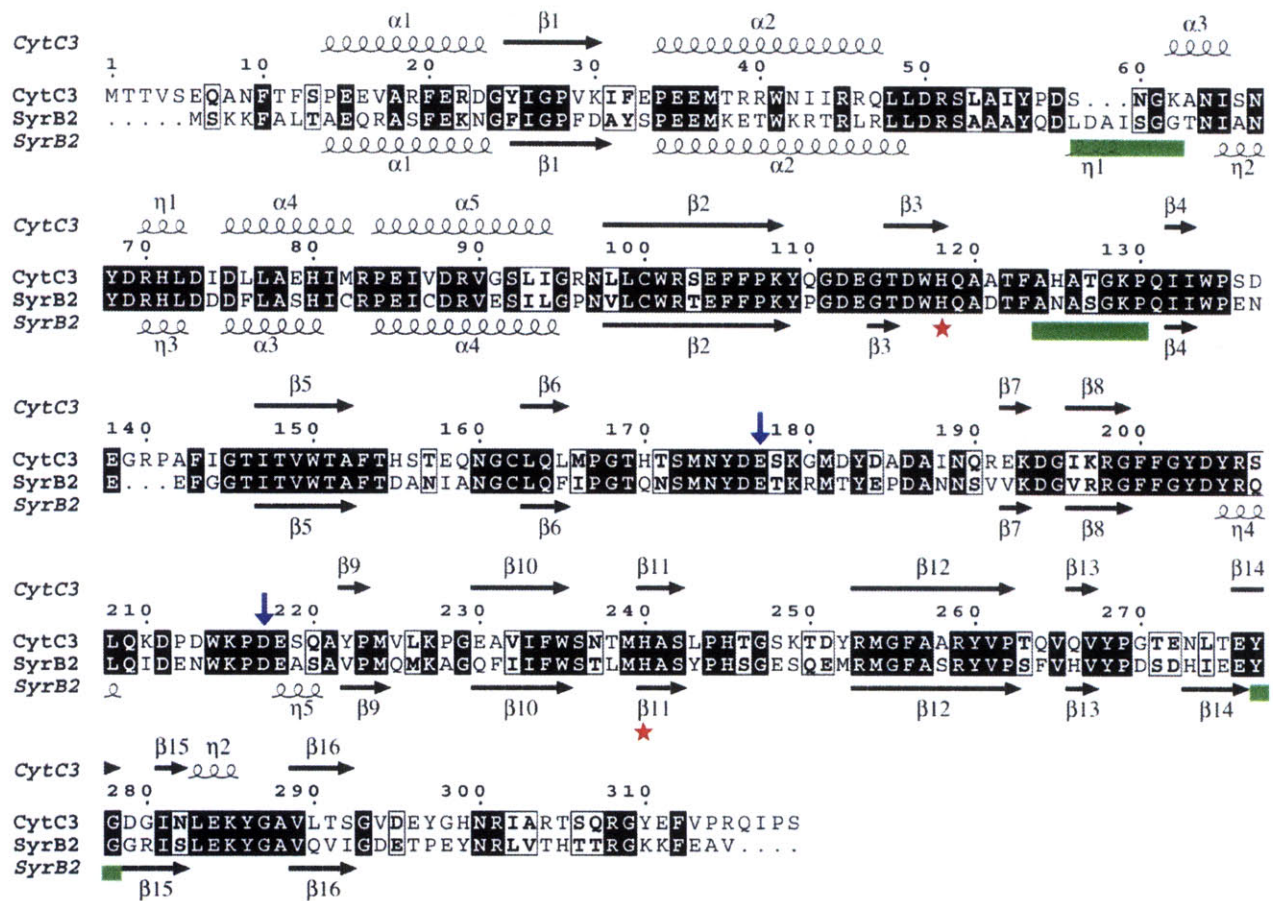
```
                    α1        β1                α2                          α3
CytC3            lllllllll   ────▶        llllllllllllll              llll
        1         10        20        30         40        50        60
CytC3   MTTVSEQANFTFSPEEVARFERDGYIGPVKIFEPEEMTRRWNIIRRQLLDRSLAIYPDS...NGKANISN
SyrB2   .....MSKKFALTAEQRASFEKNGFIGPFDAYSPEEMKETWKRTRLRLLDRSAAAYQDLDAISGGTNIAN
SyrB2            lllllllll   ────                llllllllllll         ▬▬▬▬    lll
                    α1        β1                α2                    η1      η2

          η1     α4          α5                  β2          β3            β4
CytC3    lll  lllllll  llllllllll         ──────────▶    ─────▶         ──▶
          70        80        90        100       110       120       130
CytC3   YDRHLDIDLLAEHIMRPEIVDRVGSLIGRNLLCWRSEFFPKYQGDEGTDWHQAATFAHATGKPQIIWPSD
SyrB2   YDRHLDDDFLASHICRPEICDRVESILGPNVLCWRTEFFPKYPGDEGTDWHQADTFANASGKPQIIWPEN
SyrB2    lll  llllll  llllllllll         ──────────▶   ──▶ ★ ▬▬▬▬▬▬▬▬   ──▶
          η3     α3          α4                  β2       β3            β4

               β5          β6                        β7  β8
CytC3        ─────▶     ─────▶                      ──▶ ──▶
        140       150       160       170   ▼180      190       200
CytC3   EGRPAFIGTITVWTAFTHSTEQNGCLQLMPGTHTSMNYDESKGMDYDADAINQREKDGIKRGFFGYDYRS
SyrB2   E...EFGGTITVWTAFTDANIANGCLQFIPGTQNSMNYDETKRMTYEPDANNSVVKDGVRRGFFGYDYRQ
SyrB2        ─────▶     ─────▶                      ──▶ ──▶              lll
               β5          β6                        β7  β8              η4

             β9        β10       β11             β12      β13       β14
CytC3       ──▶       ──▶       ──▶             ──────▶   ──▶       ──
        210    ▼220      230       240       250       260       270
CytC3   LQKDPDWKPDESQAYPMVLKPGEAVIFWSNTMHASLPHTGSKTDYRMGFAARYVPTQVQVYPGTENLTEY
SyrB2   LQIDENWKPDEASAVPMQMKAGQFIIFWSTLMHASYPHSGESQEMRMGFASRYVPSFVHVYPDSDHIEEY
SyrB2   l         lll       ──▶       ──▶             ──────▶   ──▶       ──▶ ▬
                  η5        β9        β10       β11   ★         β12      β13     β14

           β15  η2    β16
CytC3    ▶  ──▶ lll  ──▶
        280       290       300       310
CytC3   GDGINLEKYGAVLTSGVDEYGHNRIARTSQRGYEFVPRQIPS
SyrB2   GGRISLEKYGAVQVIGDETPEYNRLVTHTTRGKKFEAV....
SyrB2   ▬  ──▶       ──▶
           β15       β16
```

Figure IV.4. CytC3 active site region. a) Open active site in CytC3 structure allows for molecules from crystallization solution to bind, shown in stereo view. A $2F_o$-F_c composite omit map contoured at 1σ is shown in blue mesh around excess ligands (αKG and sulfate). Iron is shown in brown, and waters are shown in red. b) An iron anomalous difference map contoured at 12σ is shown in brown mesh around the iron. The αKG is labeled with the numbering nomenclature for each oxygen atoms. c) A $2F_o$-F_c composite omit map contoured at 2σ is shown in blue mesh around the active site ligands.
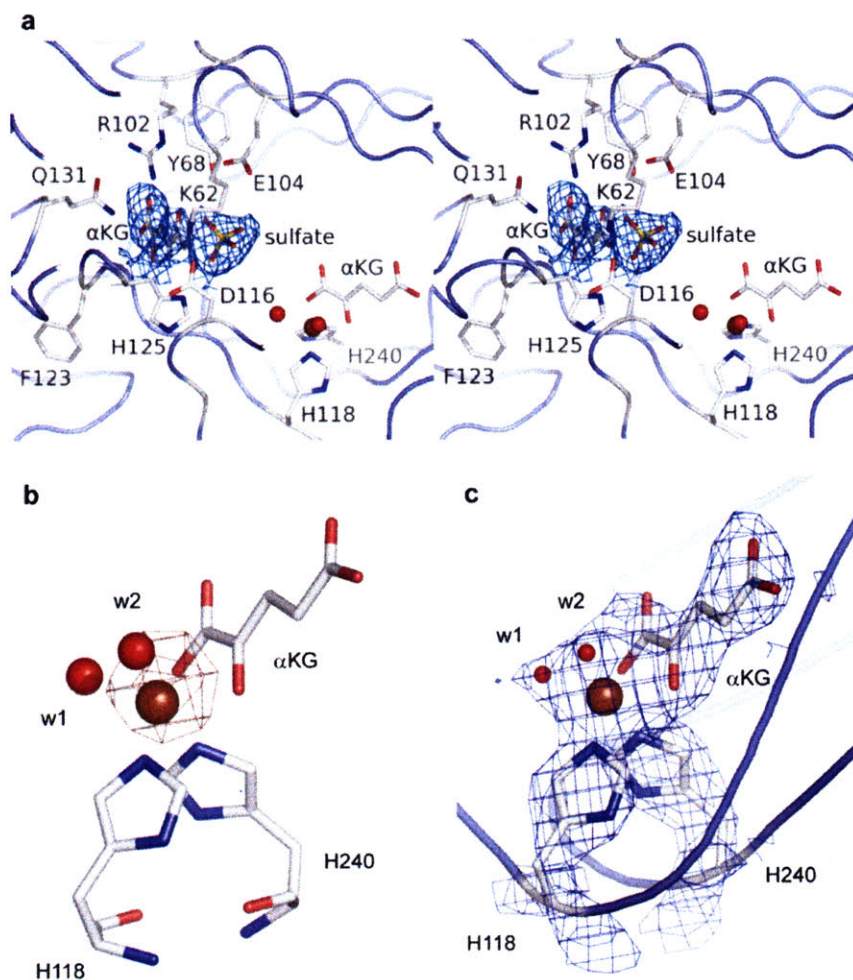
Figure IV.5. Active site comparison for CytC3 and SyrB2. Iron is shown in brown, chloride ion is shown in green, and waters are shown in red. a) Comparison between CytC3's "open" active site (blue) and SyrB2's "closed" active site (pink), where structures are aligned by F106, K108, H125, W150, T148, R253 and R259 of CytC3 with the corresponding residues in SyrB2. b) Comparison of CytC3 (blue) and SyrB2 (pink) where structures are aligned by the two active site histidines. c) Active site residues of SyrB2 and relevant distances. Dash lines indicate hydrogen bonding with labeled distances. d) Active site residues of CytC3 and relevant distances. The corresponding hydrogen bond interactions between the protein and αKG as observed in SyrB2 are indicated by black dashes. New interactions observed between CytC3 and αKG are indicated by green dashes with labeled distances.

Figure IV.5
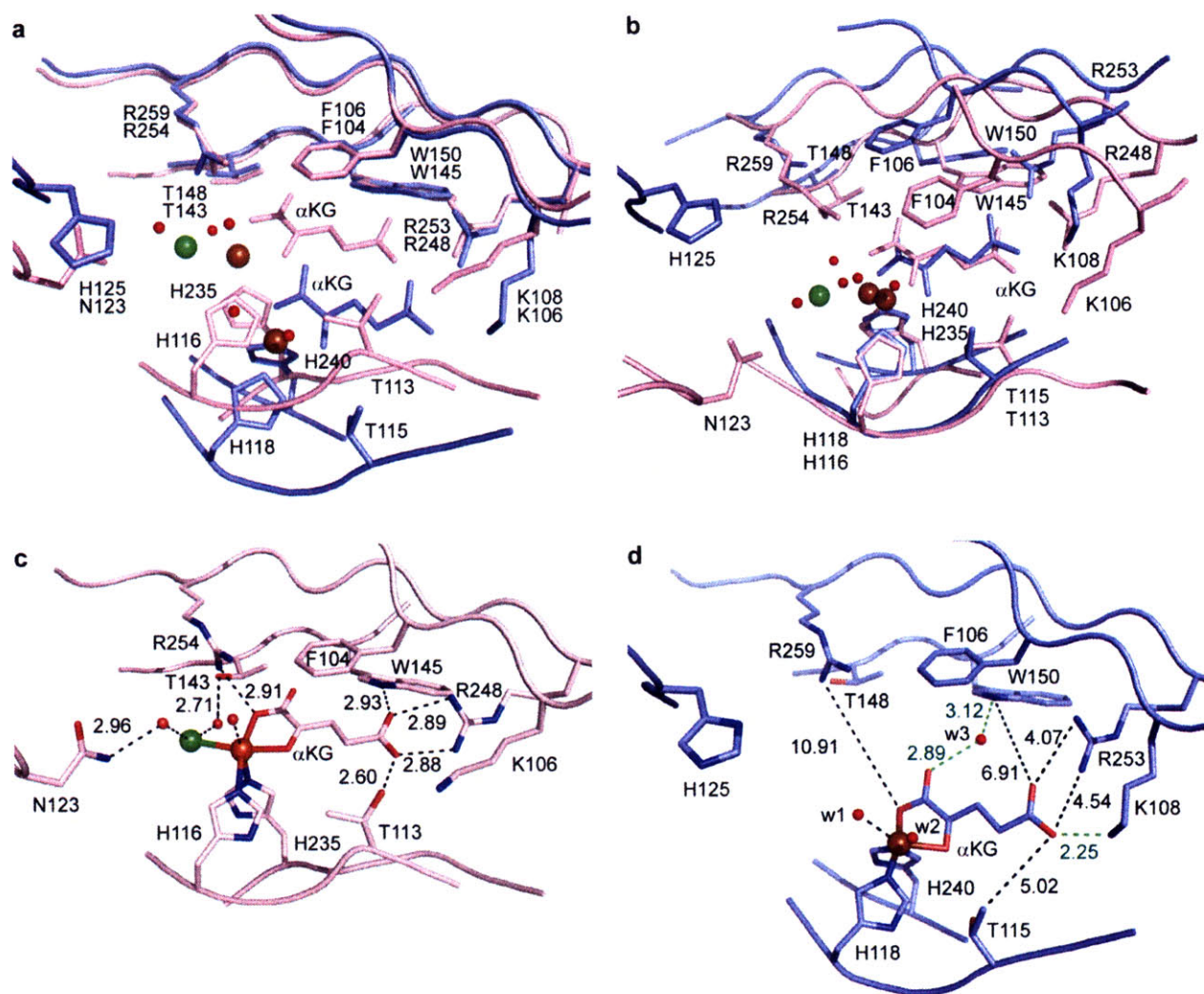
Figure IV.6. Chloride binding pocket of SyrB2 and CytC3. Iron is shown in brown, chloride ion is shown in green, and waters are shown in red. a) Residues A118, F121 and the β-carbon of S231 form a nice hydrophobic pocket for chloride in the closed state of SyrB2 active site. b) The corresponding residues, A120, F123 and S236, in the open state of the CytC3 active site are too far from the chloride binding site to form a hydrophobic pocket environment.
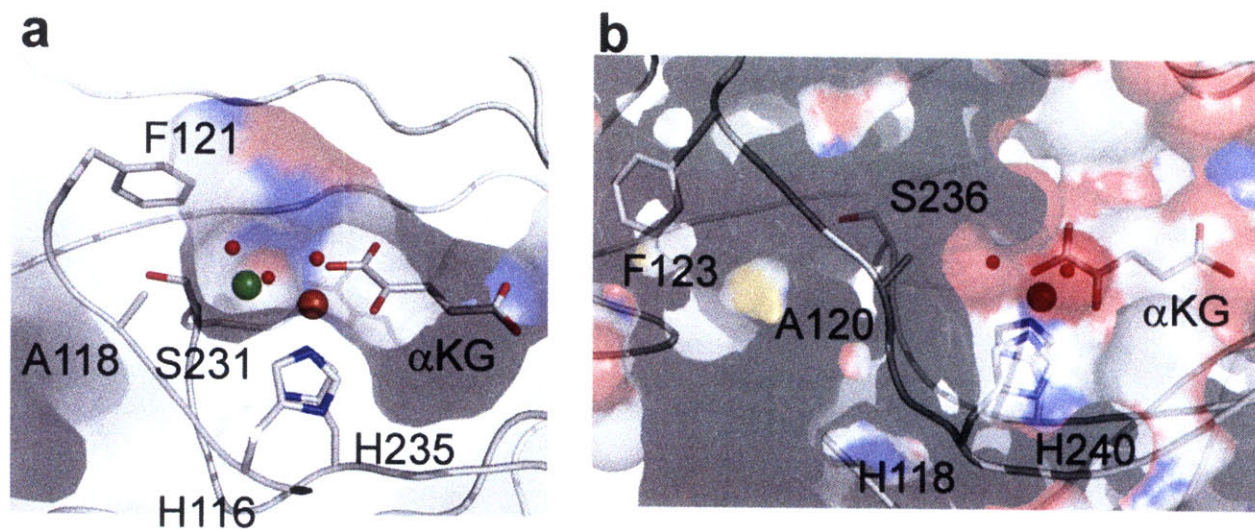
Figure IV.7. Valine amino acylated holo CytC2 substrate analogs. a) Natural CytC2 substrate. b) Analog with an amide linkage between the valine amino acid and the phosphopantetheine arm. c) Analog with an ester linkage between the valine amino acid and the phosphopantetheine arm. d) Analog with a methylene linkage between the valine amino acid and the phosphopantetheine arm.
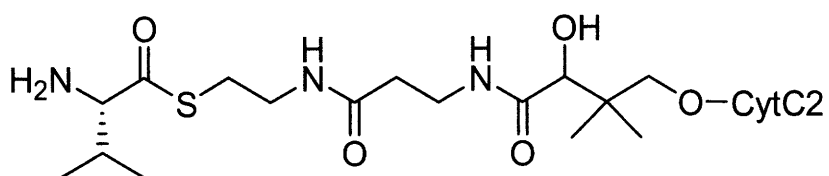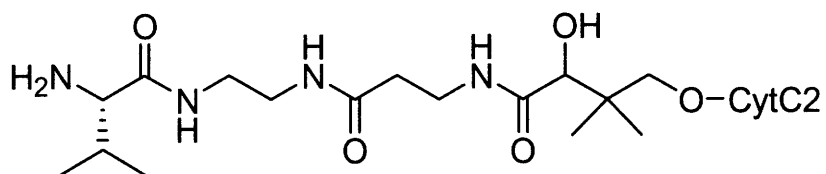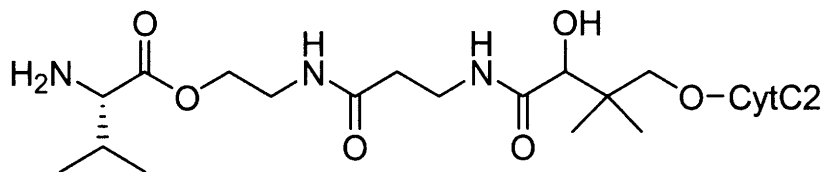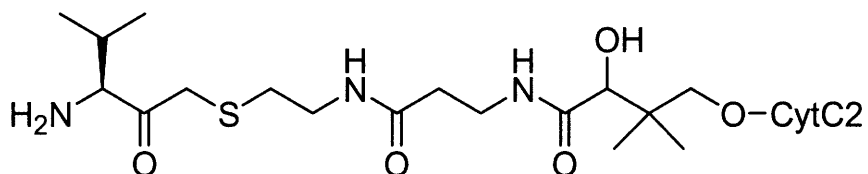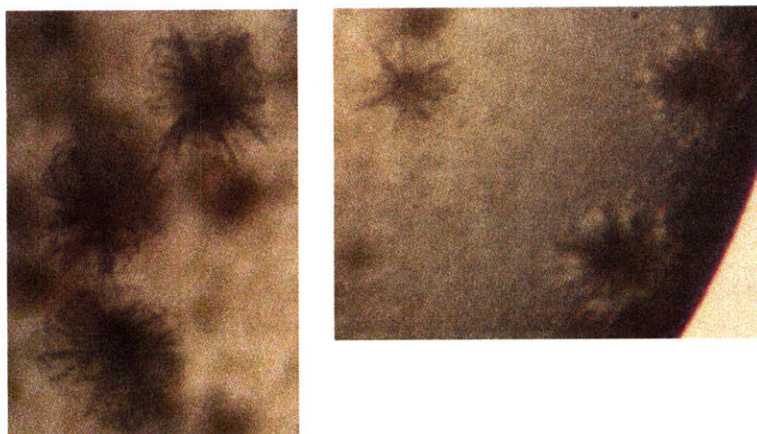
Figure IV.8. Crystals of CytC2/CytC3 complex. a) Needle haystack crystals of CytC2/CytC3 complex grown in 0.1 M Bis-Tris, pH 5.5, 0.2 M ammonium sulfate, and 30% PEG 3,350 at room temperature in Coy Chamber. b) SDS-PAGE of 3 clusters of crystals shown in a. Lane 1, Crystals were dissolved in water after three washes in mother liquor. The band corresponding to CytC3 is at the top. The theoretical molecular weight of CytC3 is 37 kDa. The band corresponding to CytC2 is denoted by the second arrow. The theoretical molecular weight of CytC2 is 12 kDa. The two bands are probably due to phosphopantetheine hydrolysis. Lane 2, molecular weight marker.
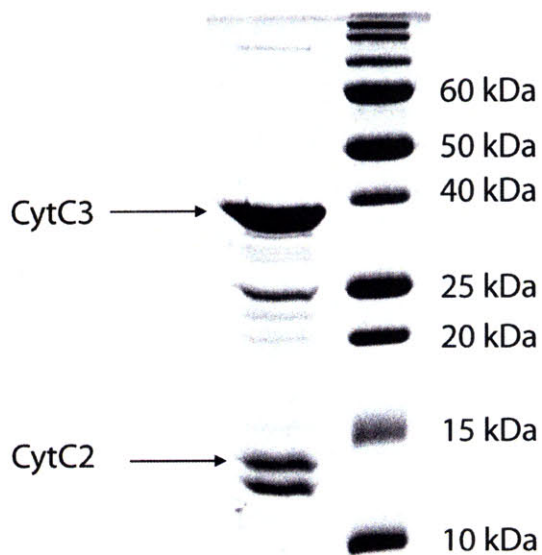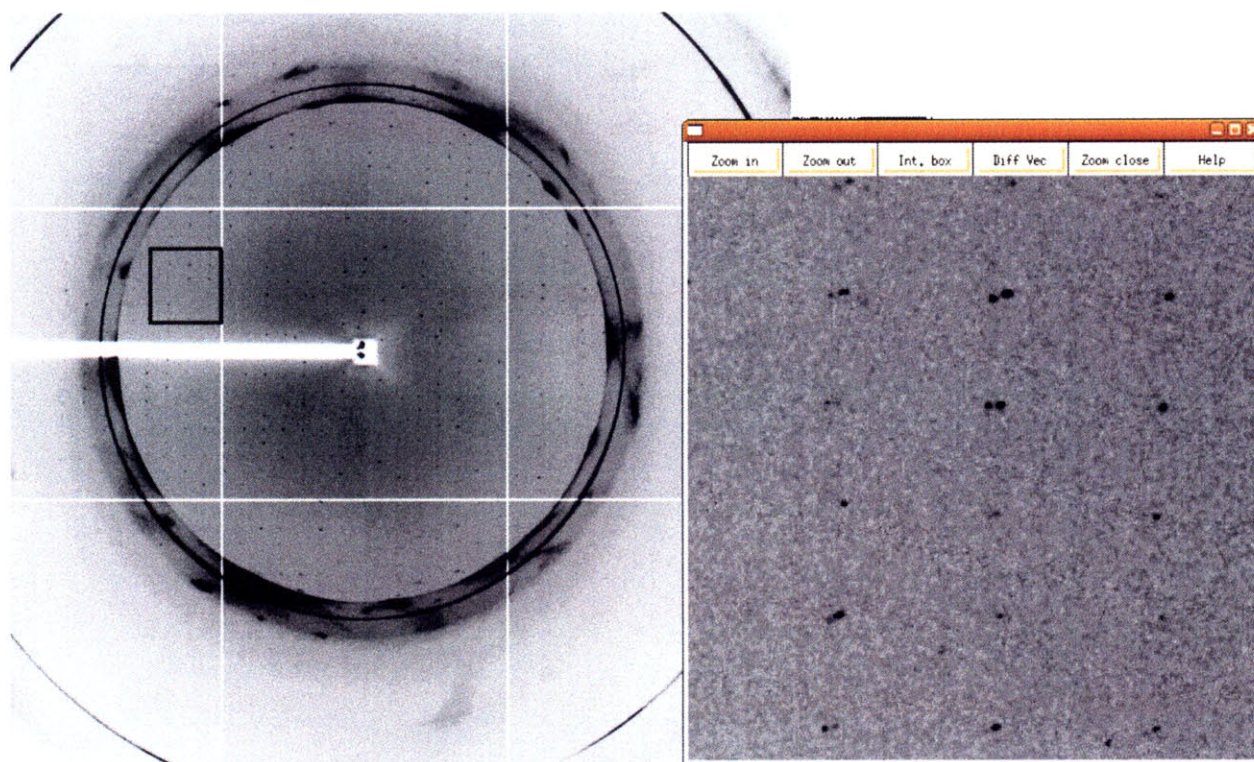
a

b

Figure IV.9. Diffraction pattern of crystal of CytC2/CytC3 complex.

Cintyu Wong

77 Mass Ave.
68-694D
Cambridge, MA 02139
(617) 258-7021
cywong@mit.edu

Education

Massachusetts Institute of Technology, Cambridge, MA
Ph.D. Candidate, Biological Chemistry. (Degree Expected May 2009)
Thesis: Crystallographic analysis of a halogenase, and a flavin dependent DNA repair enzyme.

University of Rochester, Rochester, NY
B.S., Chemistry. B.A. Economics, May 2003
Thesis: Structural Characterization of M61A Hydrogenobacter thermophilus Cytochrome C_{552}.

Research Experience

MIT, Dept. of Chemistry, Cambridge, MA
Advisor: Professor Catherine L. Drennan
The main focus is on interpreting protein structures and relating it to its functions to better understand how enzymes perform its chemistry. Work involves solving the crystal structures of a halogenating enzyme and a DNA repair enzyme. (Sep 2003 – present)

University of Rochester, Dept. of Chemistry, Rochester, NY
Advisor: Professor Kara Bren
Thesis objective was to understand how thermophilic bacteria survive under extreme temperatures. Work involves using molecular biology to create cytochrome C mutants, using techniques such as circular dichroism (CD), UV absorption spectroscopy, and nuclear magnetic resonance (NMR) to characterize the biophysical properties of the cytochrome C mutants. (Sep 2000 – May 2003)

Schering Plough Pharmaceutical, Dept. of Protein Science, Kenilworth, NJ
Advisor: Dr. Hung Le
Project objective was to investigate the enzymatic activity of *Hepatitis C* virus (HCV) mutants. Much work involved protein expression and purification of HCV, and performed ^{32}P radioactive assays. (Summer 2002)

University of Notre Dame, Dept. of Chemical Engineering, Notre Dame, IN
Advisor: Professor Agnes E. Ostafin
Project objective was to develop an immuno-biosensor for detecting viruses, and to find a mechanism to deliver drugs/enzymes to infected areas using

silica nano-shells. Modified the surface chemistry of silica nano-shell for immunoassay, and used ELISA assay to quantify the efficiency of antibody attachment to silica nano-shells. (Summer 2001)

University of Pennsylvania, Dept. of Biochemistry & Biophysics, Philadelphia, PA
Advisor: Professor Paul Liebman
Project objective was to investigate the effects of different membrane phospholipids composition on metarhodopsin formation. Used dynamic light scattering to examine the homogeneity of rhodopsin under different conditions. (Summer 2001)

Teaching Experience

MIT, Department of Chemistry, Cambridge, MA
Biophysical Chemistry Techniques Teaching Assistant
Taught and led workshops of ~20 students in graduate level biophysical course. Class concentrated on both the theory and practical use of X-ray crystallography in determining the 3-D structure of macromolecules. (Jan 08 – May 08, Jan 05 – May 05)

HHMI-MIT Summer Research Program Mentor
Trained and guided undergraduate student to complete a research project. Taught basic biochemistry skills including protein purification and crystallization techniques. Worked as an advisor for the individual's project. (Summer 2008)

General Chemistry Teaching Assistant
Taught two recitation lectures to ~25 students each week, as well as tutored students one-on-one during office hours. Class introduced the fundamentals of Chemistry including thermodynamics, chemical reactions and equilibria, and molecular orbital theory. (Sep 03 – Dec 03)

University of Rochester, Department of Chemistry, Rochester, NY
Inorganic Chemistry Teaching Assistant
Held small discussion groups during class, and facilitated problem solving in a group. The class focused on molecular orbitals, point theory, and properties of transition metals. (Sep 02 – Dec 02)

Organic Chemistry I & II Workshop Leader
Led small group of ~10 students in a workshop format. Assisted students individually with problem sets. The workshop used an observation-deduction technique to help students to better understand chemical reactions, and to easily work through problem sets. (Sep 00 – May 01)

Publications

C. Wong, D. P. Galonic, C.T. Walsh, and C. L Drennan, Structural Analysis of the Open Active Site Conformation of CytC3: an Alpha Ketoglutarate Nonheme Iron Halogenase. *JACS* (2009), 131(13), 4872-79.

L. E. Frick, J. C. Delaney, <u>C. Wong</u>, C. L.Drennan, J. M. Essigmann, Alleviation of 1,N^6-Ethanoadenine Genotoxicity By the *Escherichia Coli* Adaptive Response Protein Alkb; *PNAS* (2007), 104(3), 755-60.

J. C. Delaney, L. Smeester, <u>C. Wong</u>, L. E. Frick, K. Taghizadeh, J. S. Wishnok, C. L. Drennan, L. D. Samson, J. M. Essigmann, Alkb Reverses Etheno DNA Lesions Caused By Lipid Oxidation *In Vitro* and *In Vivo*; *Nat. Struct. Mol. Biol.* (2005), 12(10), 855-60.

Presentations   *Nature's Way to Halogenation: an Environmentally Friendly Alternative,* <u>C. Wong</u>, D. P. Galonic, C. T. Walsh, C. L. Drennan; 2008 Synethic Biology Conference 4.0 in Hong Kong, poster presentation.

Crystal Structure of the Open Active Site Conformation of CytC3, <u>C. Wong</u>, D. P. Galonic, C. T. Walsh, C. L. Drennan; Biology and Chemistry Inter-Departmental Seminar at MIT, MA, oral presentation.

Crystal Structure of an Alpha Ketoglutarate Nonheme Iron Halogenase CytC3, <u>C. Wong</u>, D. P. Galonic, C. T.Walsh, C. L. Drennan; 2007 234th ACS National Meeting in Boston, MA, poster presentation.

Crystal Structure of an Alpha Ketoglutarate Nonheme Iron Halogenase CytC3, <u>C. Wong</u>, D. P. Galonic, C. T. Walsh, C. L. Drennan; 2007 25th Summer Symposium in Molecular Biology "Frontiers in Metallobiochemistry" at Pennsylvania State College, PA, poster presentation.

Awards   American Chemical Society Award for Excellence in Chemistry, Rochester (2003)

Willard Line Scholarship, Rochester (2001 – 2003)

University of Rochester Chemistry Department Award, Rochester (2002).

Activities   Member, American Chemical Society (2007 – present)

Board Member, MIT Chinese Student Scholar Association (2007 – present)

Manager, Ashdown Graduate Housing (2005 – present)

Officer, Ashdown Graduate Housing (2005)