

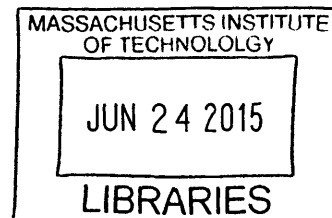
Structural Studies of Radical Enzymes in Bacterial Central Metabolism

by

Michael A. Funk

B.S., Chemistry, Biology (2008)
Vanderbilt University

ARCHIVES



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

June 2015

© 2015 Massachusetts Institute of Technology
All Rights Reserved

Signature redacted

Signature of Author _____

Department of Chemistry
May 8, 2015

Signature redacted

Certified by _____

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

Signature redacted

Accepted by _____

Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Studies

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

Professor Barbara Imperiali Signature redacted
Committee Chair
Class of 1922 Professor of Chemistry and Biology

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

Professor JoAnne Stubbe Signature redacted
Committee Member
Novartis Professor of Chemistry and Professor of Biology

Structural Studies of Radical Enzymes in Bacterial Central Metabolism

by

Michael A. Funk

Submitted to the Department of Chemistry on May 8, 2015
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biological Chemistry

ABSTRACT

Anaerobic bacteria play a crucial role in cycling of nutrients in diverse ecosystems, degradation of organic compounds, and as key members of the human gut microbiome. The absence of oxygen limits the chemistry that bacteria can perform; however, these organisms do make use of organic radical cofactors that are oxygen sensitive. This thesis presents a structural analysis of three enzymes that utilize a glycyl radical cofactor to perform difficult, radical-based chemistry. Benzylsuccinate synthase catalyzes the first step in the anaerobic degradation of toluene, a major component of gasoline and an environmental pollutant. Choline trimethylamine-lyase is used by gut bacteria to degrade choline, producing a byproduct, trimethylamine, which is linked to human diseases. These two enzymes share a common protein fold and also utilize a similar series of steps to initiate chemistry on their substrates. However, once they have generated a radical intermediate, the enzyme active site guides very different chemical steps in these two enzymes. Class III ribonucleotide reductases catalyze the same chemical reaction as observed in aerobic or oxygen independent systems, but utilize a glycyl radical cofactor to initiate chemistry. X-ray crystal structures have given us snapshots of these enzymes in action and accompanying biochemical experiments are help reveal the mechanisms they use to control radical chemistry. These structures have highlighted the potential diversity of chemical reactions available to anaerobic organisms through the use of a conserved enzyme fold.

Thesis Supervisor: Catherine L. Drennan

Title: Professor of Chemistry and Biology
Howard Hughes Medical Institute Investigator and Professor

DEDICATION

This thesis is dedicated to my mother, Margie, the very first in a long series of tough, generous, inspiring, and demanding women who have helped me through my scientific journey to this point.

ACKNOWLEDGEMENTS

I first want to thank my advisor, Cathy Drennan, for both giving me the opportunity to find my own path and for holding me to high standards in such a positive and motivating way. My thesis may be an academic and intellectual composition, but the skills and wisdom I have gained from working with Cathy are so much more than what can be found in this document. In addition to being one of the sharpest scientists I know, Cathy is a genuinely kind person who cares deeply about her students. I am so thankful to have been able to be a part of the wonderful environment that is the Drennan lab at MIT.

I am so very thankful to have had wonderful collaborators and advisors throughout my graduate career. My thesis chair, Barbara Imperiali, has been a source of inspiration and encouragement since I first arrived at MIT and was part of the ChemBio Interface program she lead. I am so fortunate to have had many interactions with JoAnne Stubbe, whether as a 5.07 TA, a rotation student in her lab, a collaborator on several projects, or merely chatting with her at department events. Whether directly or indirectly, she has made me think harder and put more effort into almost everything I have done at MIT, for which I am so grateful.

My way to MIT was paved by my mentor at Vanderbilt, Tina Iverson. I cannot begin to express my gratitude to her for taking me in and providing me with so many opportunities. As the second person to start working in the Iverson lab, I was trained directly by Tina and was able to participate in the lab at a level I could not have imagined before. Tina selflessly encouraged me to apply for a summer position at Caltech my junior year, helped me make connections with Cathy before I applied to MIT, and made sure I applied for the NSF Graduate Research Fellowship my senior year. She was instrumental in so many early successes that have helped me arrive where I am now, and she continues to be a source of inspiration for me as I aspire to continue in academic research and teaching.

My collaborators in all my projects have been wonderful sources of ideas and discussion, and my thesis would not be half as complete without their help and support. In this regard, I must first acknowledge all the members of the Drennan lab, but especially Ed, Nozomi, and Christina with whom I worked closely for years and shared the joy, pain, and—so often—utter confusion that is scientific discovery. Yifeng Wei has been an amazing source of inspiration, discussion, encouragement, and—of course—raw materials for my research. I can't begin to describe how thankful I am that he kept coming to me when he had a new and exciting idea. At the end of my 4th year, I began a collaboration with Smaranda Bodea and Emily Balskus that has filled up the rest of my time in the Drennan lab with great success. Sam's constant support both in terms of materials and discussions has been amazing.

I owe so much of my success at MIT to some of my fellow Biological Chemistry first years—Jingnan, Vinita, Yifeng, Marco, Ken, Hongik, and Tengfei. We helped each other through tutorial, cumes, and orals, and we still manage to go out to Mulan periodically. Beyond

academics, I cannot begin to acknowledge Jingnan and Christina—they have has been there supporting me through wonderful, scary, and sad moments.

I'm so thankful for everyone in the Drennan lab. Your lab becomes a bit like family over the years—you come to care deeply for the people around you. They are an amazing group of smart, caring, and fun people. Over the years I've a few kindred spirits in the lab, and to them I am especially indebted: for keeping me sane, humble, and laughing. In particular, Jen and Tsehai have been my support network during the writing and editing of this thesis, and Yang-Ting has provided generous, and well-meant, skepticism.

Finally, I am so thankful for the constant loving support of two people throughout this journey: my mother, Margie, and my dear friend, Lisa. My mother has from my very earliest years instilled and encouraged a deep curiosity in me about the natural world. I am so happy that as we have both grown older that link has strengthened and now we can engage each other in myriad fun discussions of science, society, and nature. I think of Lisa more as a sister than a friend. She has helped me stay sane through these six years, listening to my frustrations and keeping me down to earth when I needed it.

TABLE OF CONTENTS

Chapter I: Glycyl radicals in bacterial central metabolism	13
I.A. Summary	14
I.B. The impact of oxygen on bacterial metabolism	14
I.C. Anaerobic metabolism in bacteria	17
I.D. Alternatives to oxygen in anaerobes	20
I.E. Glycyl radical enzymes are key players in anaerobic metabolism	24
I.F. Core features of glycyl radical enzyme structure and catalysis	28
I.G. Glycyl radical enzyme activation by activating enzymes	32
I.H. Goals of the thesis	39
I.I. References	41
Chapter II: Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity	47
II.A. Summary	48
II.B. Introduction	48
II.C. Results	52
<i>Overall structure of BSS</i>	52
<i>BSS inter-subunit interactions</i>	61
<i>Global structural changes in the BSS $\alpha\gamma$ complex</i>	62
<i>Limited proteolysis of BSS complexes</i>	65
II.D. Discussion	65
II.E. Materials and Methods	73
II.F. Acknowledgements	78
II.G. References	78
Chapter III: The structure of a pre-turnover complex of benzylsuccinate synthase guides mechanistic understanding of the X-succinate synthase enzyme family	83
III.A. Summary	84
III.B. Introduction	84
III.C. Results	87
<i>Substrates can bind to BSS$\alpha\beta\gamma$</i>	87
<i>Fumarate binding stabilizes a partially closed state of BSS$\alpha\gamma$.</i>	89
<i>BSS lines up fumarate and toluene for radical chemistry</i>	94
<i>Two active site regions contribute to specificity in X-succinate synthases.</i>	97
III.D. Discussion	102
III.E. Materials and Methods	107
III.F. Acknowledgements	111
III.G. References	111

Chapter IV: Structural and biochemical characterization of choline trimethylamine-lyase suggests how a glycyl radical enzyme binds and cleaves choline	117
IV.A. Summary	118
IV.B. Introduction	118
IV.C. Results	123
<i>Choline TMA-lyase adopts a conserved GRE fold.</i>	123
<i>Choline interacts with polar residues within the CutC active site.</i>	125
<i>Comparison of the CutC structure with its homology model reveals the difficulty in predicting active site features in GREs.</i>	128
<i>Glycyl radical installation on wild type and mutant CutC proteins is variable.</i>	130
<i>Mutations of tyrosines involved in CH–O interactions yield enzyme that is still active.</i>	133
<i>Mutation of residues in conserved GRE motifs suggest conservation of function as well as sequence.</i>	134
<i>A putative proton transfer network in the active site of CutC.</i>	136
IV.D. Discussion	138
IV.E. Materials and Methods	146
IV.F. Acknowledgements	153
IV.G. References	153
Chapter V: The class III ribonucleotide reductase from <i>Thermotoga maritima</i> can utilize thioredoxin as a reductant	163
V.A. Summary	
V.B. Introduction	
V.C. Results	170
<i>NbNrdD Is a Glycyl radical Enzyme.</i>	170
<i>NbNrdD Catalyzes CTP Reduction Using TrxA/TrxB/NADPH.</i>	171
<i>NbNrdD(C301A) Is Inactive, and Reaction of NbNrdD(C300A) with CTP Generates Cytosine.</i>	173
<i>Crystal Structure of TmNrdD Allows Modeling of Active Site Residues in Redoxin-Dependent NrdDs.</i>	175
<i>Bioinformatics Analysis Suggests Three Chemically Distinct NrdD Subtypes.</i>	182
V.D. Discussion	183
<i>Biochemistry, Structure, and Bioinformatics Support the Existence of a Second NrdD Subtype.</i>	183
<i>Distribution of NrdD Subtypes Correlates with Metabolism.</i>	184
<i>A Clue Regarding the Ancestral NrdD.</i>	186
V.E. Materials and Methods	187
V.F. Acknowledgements	196
V.G. References	197

Chapter VI: Recognizing the diversity within the glycy radical enzyme family	205
VI.A Summary	206
VI.B Unexplored function of glycy radical enzyme family members	206
VI.C The phylogenic distribution of GREs and GRE metabolic steps.	216
VI.D The special case of pyruvate formate-lyase	219
VI.E The origin of GREs	221
VI.F An explanation for the evolution of glycy radical enzymes	222
VI.G The distribution and evolutionary history of activating enzymes	223
VI.H Conclusion	226
VI.I Acknowledgements	227
VI.J References	227
Curriculum vitae	231

Chapter I

Glycyl radical enzymes in bacterial central metabolism

I.A Summary

Bacteria are arguably Nature's most accomplished opportunists. Bacterial communities thrive in virtually every environment on Earth where there is some source of chemical or photochemical energy to be harnessed. Rich environments, such as the guts of higher eukaryotes, provide ample sugars and amino acids to support a simple, fermentative metabolism, but nutrient-poor spaces, such as the open ocean, require complex chemical transformations to handle the fundamental elements of life. Unique adaptations have enabled bacteria to thrive within the most mundane and the most extreme environments in ways that are often unusual and counterintuitive. In particular, molecular oxygen, and the lack thereof, enables or prohibits the chemical processes available to bacteria. Within Chapter I, I will explore the diversity of bacterial anaerobic metabolism and define the roles of a particular class of strictly anaerobic biological catalysts, the glyceryl radical enzymes (GREs). GREs use a glyceryl radical cofactor to catalyze a diverse range of chemistry, some of which is unique and some of which occurs through other means under aerobic conditions. GREs share a common fold and activation mechanism, and some enzymes share putative radical intermediates. These enzymes contribute at multiple levels to anaerobic metabolism and play an important role in bacterial metabolism, bioremediation, and virulence.

I.B. The impact of molecular oxygen on bacterial metabolism

One of the most profound environmental factors living organisms face is the presence or absence of oxygen. Molecular oxygen (O_2) has an unusual electronic configuration in that its ground state form is a triplet diradical, having two unpaired electrons occupying π^* antibonding orbitals (Figure I.1A) (1). This electronic structure makes O_2 a stable component of the atmosphere, as the unpaired electrons do not readily react with molecules with paired electrons. However, O_2 does react irreversibly with organic radicals (Figure I.1B), molecules containing an unpaired electron. The resulting peroxy compounds are highly oxidizing and can damage nearby molecules to yield organic peroxides. Likewise, reduced iron-sulfur clusters, flavins and some other redox cofactors can be oxidized by O_2 to yield superoxide and/or hydrogen peroxide (2, 3). On their own, these reduced oxygen species are not particularly reactive, although they can damage redox cofactors like iron-sulfur clusters and protein cysteine residues. Bacteria living in the presence of O_2 contain protective enzymes and small molecule antioxidants that rapidly

eliminate these reduced oxygen species before they can accumulate or react to form highly reactive species, the most notorious of which is the hydroxyl radical, which can do damage to proteins, lipids, or DNA (1). In aerobic organisms, enzymes that utilize organic radicals or some redox-active cofactors must be protected from O₂ by the protein scaffold or repressed in the presence of O₂.

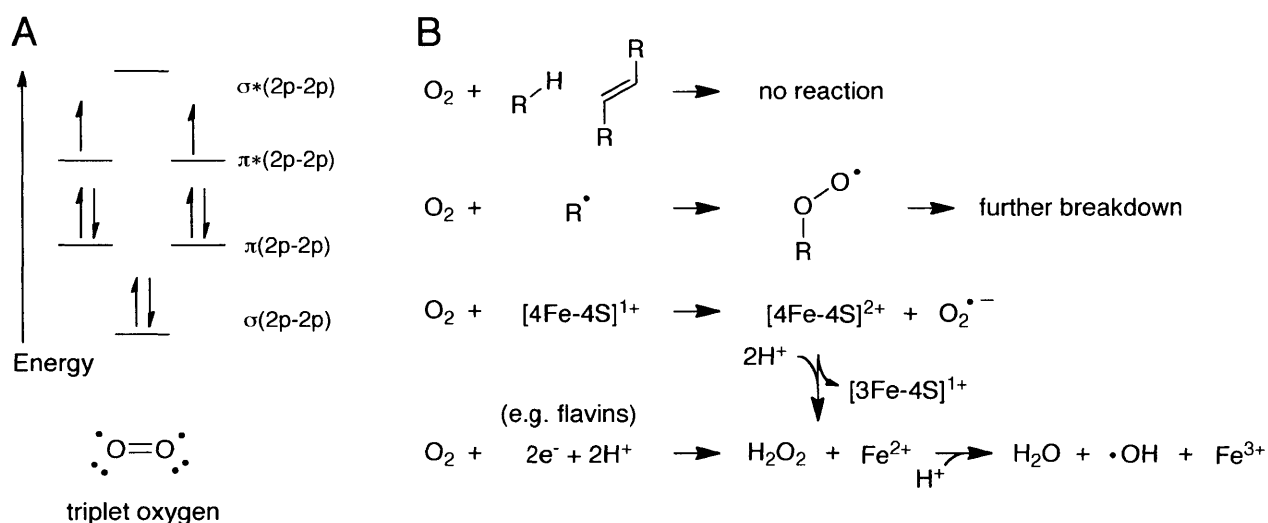


Figure I.1. The electronic structure of oxygen creates challenges for radical enzymes. (A) Molecular orbital diagram of the 2p orbitals of molecular oxygen (O₂) showing the triplet diradical structure. **(B)** O₂ reacts to varying degrees with different compounds. Most non-radical organic compounds are inert to O₂; however, organic radicals react rapidly and irreversibly to form organic peroxy compounds that can form secondary radical products. Reduced, low-potential iron-sulfur clusters are especially sensitive as they may be oxidized by O₂ to form superoxide (O₂^{•-}). Further cluster degradation and oxidation yields hydrogen peroxide (H₂O₂) and free Fe²⁺, a particular noxious combination that can yield hydroxyl radicals (HO•) through Fenton chemistry. H₂O₂ can also form as a result of 2 e⁻ reduction by redox-active cofactors such as flavins. H₂O₂ and O₂^{•-} are not highly reactive themselves, but their accumulation is toxic as subsequent reactions can yield products such as HO• that are extremely reactive and damaging to biomolecules. A host of enzymes and small molecules are present in organisms that are oxygen tolerant to eliminate these compounds and mitigate damage.

Despite these challenges, O₂ also provides a huge advantage to organisms: it serves as a convenient, 4 e⁻ oxidant for electron transport enzymes during respiration (discussed below). In this process, electrons from organic molecules are passed through a series of organic and metal redox cofactors of higher and higher reduction potential. Along the way, the energy produced by this downhill movement of electrons is used to pump protons across a membrane, generating an electrochemical gradient that provides energy for membrane transport, flagellar motion in

bacteria, and ATP synthesis (2). Anaerobic respiration also occurs with sulfate, nitrate, oxidized metal or organic ions, or halogenated compounds as electron acceptors, but these common electron acceptors are not as oxidizing as O₂ and thus yield less potential energy. O₂ can also support oxidative chemistry outside of respiration; for example, maturation of cofactors such as heme requires O₂ in many organisms (4), as described below.

One way to classify bacterial metabolism is with regard to tolerance of O₂. In all aerobic organisms, cellular antioxidants such as glutathione and enzymes such as catalase, superoxide dismutase, and peroxidases provide protection from reduced oxygen species (5). However, many organisms that live in strictly anaerobic environments do not have mechanisms to protect themselves from oxidative stress. Without these safeguards, O₂ can wreck havoc through direct oxidation of sensitive compounds or cofactors, and reduced oxygen species can accumulate to toxic levels. Bacteria can be classified based on their requirements for O₂ (5): For **obligate aerobes**, adequate O₂ for aerobic respiration is a growth requirement. **Microaerophiles** typically require O₂ to perform essential biosynthetic reactions, but do not use O₂ for respiration, and thus prefer low O₂ concentrations found in various habitats away from direct contact with air. Many pathogens require low O₂ levels. **Aerotolerant** organisms may protect themselves from O₂, but do not use it in their metabolism in any way and are thus suited to anaerobic environments that may have periodic influxes of O₂ (5). **Facultative anaerobes** can use O₂ when it is available, but can also grow through fermentation or respiration with alternate electron acceptors. They typically have both aerobic and anaerobic pathways for biosynthesis of essential cofactors. **Obligate anaerobes** typically do not have the protective measures mentioned above and are thus susceptible to even small amounts of O₂. A large component of this sensitivity is that many obligate anaerobes utilize enzymes and cofactors that are damaged, often irreversibly, by O₂.

The phylogenetic distribution of anaerobic and aerobic organisms is complex. There is some consensus that the early Earth lacked abundant O₂ (6, 7). Thus it is presumed that the first metabolisms were strictly anaerobic with little protection from O₂ and its reduction products. The geochemical record is consistent with very low O₂ concentrations prior to the evolution of cyanobacterial oxygenic photosynthesis ~2.7 billion years ago, and subsequent widespread oxygenation did not occur until ~2.3 billion years ago, after geological redox buffers were exhausted (6). Most bacterial phyla presumably diverged before the widespread prevalence of O₂-rich environments, but O₂ tolerance can vary at the species level (8), and some organisms

presumed to be quite ancient—*Aquifex* species are predicted to have diverged 3.5 billion years ago, around the same time as the bacteria/archaea split (9)—are capable, or even obliged, to use O₂ in their primary metabolism, whereas many more recently diverged bacterial species, Gram-positive clostridia, for example, are obligate anaerobes. The explanation for this confounding pattern is likely a combination of an ancient origin for O₂ utilization and protection (10) and genetic and metabolic remodeling of bacterial species during evolution by lateral gene transfer (LGT) and gene loss (11). Indeed, aerobic respiratory complexes have been traced back to the common ancestor of bacteria and archaea (10) suggesting a very ancient origin of energy-producing O₂ reduction. The circumstances under which such a metabolism could operate within the paradigm of an anoxic early atmosphere remains a source of controversy, although recent experiments showing much lower K_M values for O₂ for some bacterial terminal oxidases suggest that such metabolisms might be feasible even in the absence of widespread oxygenation (12). Fully aerobic metabolism would have provided a huge benefit in terms of energy production, and thus permitted the huge expansion of bacteria and eukaryotes into new habitats. At the same time, strictly anaerobic processes such as methanogenesis (by archaea), acetogenesis, some kinds of fermentation, and most sulfate reduction were confined to specific habitats away from direct contact with the atmosphere or in conjunction with aerobic bacteria that could consume O₂ faster than it is introduced (8, 13).

I.C. Anaerobic metabolism in bacteria

In the absence of O₂, bacteria perform several types of metabolism to produce ATP, generate reducing equivalents, fix carbon, and synthesize essential biomolecules. In general, organisms are divided into **autotrophs**, which fix CO₂ to generate biomass, and **heterotrophs**, which scavenge fixed carbon from the environment (Figure I.2A). Likewise, divisions can be made based on source of ATP: **phototrophs** generate ATP through the reactions of photosynthesis; **lithotrophs** use various reactions involving inorganic compounds such as metal ions, CO₂, and H₂—either with or without organic compounds—to generate ATP; and **organotrophs** rely exclusively on organic compounds to generate ATP. Reducing equivalents in each of these divisions parallels ATP production: anaerobic phototrophs typically oxidize reduced sulfur or metal ions using light to boost the reducing power of the harvested electrons, lithotrophs use hydrogen, sulfide, or ammonia as a source of reducing equivalents, and

organotrophs harvest reduced compounds from their environment. Some bacteria can utilize multiple metabolic schemes and some kinds of metabolism do not fit neatly within this classification; the diversity of metabolism present in Nature defies simple categorization.

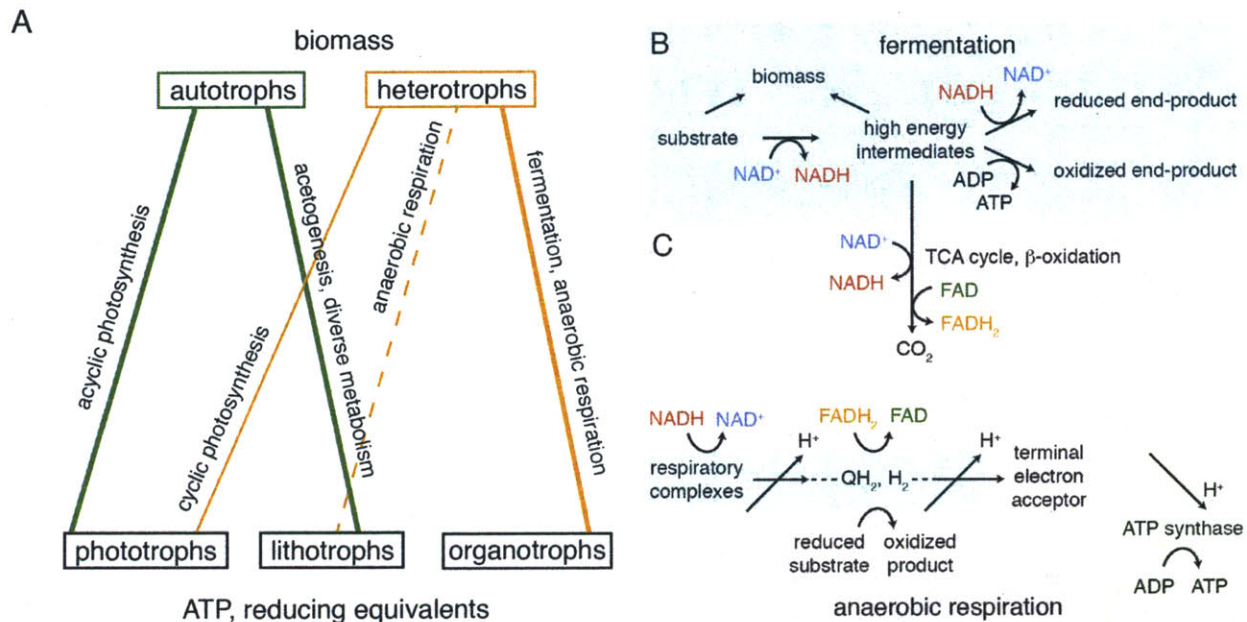


Figure I.2. Modes and mechanisms of bacterial anaerobic metabolism. (A) Bacteria can be divided into categories by their source of carbon for biomass generation and source of ATP and reducing equivalents. **(B)** Fermentation generates energy through disproportionation of substrates to more and less oxidized products. Substrate-level phosphorylation generates ATP, and some of the end products can be diverted to biomass or exported as waste. **(C)** Anaerobic respiration continues the oxidation of intermediates from fermentation (or glycolysis) pathways. Electron carriers such as FAD and NADH are utilized to fully oxidize substrates. These carriers are subsequently oxidized in membrane-imbedded respiratory complexes, which permit the formation of a proton (H⁺) gradient, typically through the use of intermediate electron carriers such as hydroquinones (QH₂) or hydrogen gas (H₂). Reduction of a terminal electron acceptor provides the driving force for this electron transport chain. Release of the proton gradient is coupled to ATP generation by an ATP synthase complex. This is a simple schematic diagram; anaerobic respiration is extremely diverse, and not all steps are depicted, nor are all steps necessarily required.

This thesis will focus on heterotrophic metabolism, specifically that of anaerobic organotrophs. As described above, these bacteria utilize fixed carbon sources for both ATP production and reducing equivalents through two major metabolic schemes (Figure I.2B-C) (2, 8):

Fermentation is the simultaneous oxidation and reduction of organic substrates coupled to substrate-level phosphorylation. Fermentations are preferably carried out with substrates with an oxidation state close to 0; sugars, amino acids, and nucleobases thus are all compatible to varying degrees. Oxidation of these compounds typically generates CoA thioesters that are converted to ATP through a phosphoanhydride intermediate, although some substrate level phosphorylation occurs without a thioester intermediate. The reducing equivalents from the oxidation step are then used to reduce the same substrates or their intermediates in order to regenerate electron carriers such as NAD^+ , and FAD. Some fermentative bacteria are able to dispose of electrons through reduction of protons to H_2 in addition to forming reduced carbon compounds. Primary fermentation products such as lactate, acetate, and ethanol are commonly subject to secondary fermentation by specialized organisms that produce longer chain organic acids. Fermentation as mode of energy conservation is wasteful in the sense that large amounts of reduced organic products are excreted without further utilization. However, in environments such as the gut, soil, and decaying organic matter where high-energy substrates are abundant and/or exogenous electron acceptors are scarce, fermentation prevails as the most efficient strategy for rapid growth for many bacteria (14).

Anaerobic respiration, at its simplest, describes the coupling of electron transfer down an electrochemical gradient within membrane bound respiratory complexes to the translocation of protons across a membrane. Electrons, typically in the form of NADH or FADH_2 , are generated from reduced organic substrates (or some inorganic substrates) through the actions of glycolysis, the tricarboxylic acid cycle, β oxidation, or other oxidative processes. Terminal electron acceptors such as oxidized metal ions (e.g. Fe(III) , Mn(IV)), inorganic oxides (e.g. NO_3^- , SO_4^{2-} , AsO_4^-), fumarate, trimethylamine-*N*-oxide or S(0) provide the driving force for moving electrons through one or more membrane-bound respiratory enzyme complexes. Protons are pumped across the membrane during electron transfer and the resulting gradient is used to power ATP synthesis by ATP synthase. In some cases, respiration goes through H_2 as an electron carrier rather than the more familiar quinones and cytochromes but always results in the generation of a proton gradient. Most facultative anaerobes use alternate electron acceptors only when O_2 is absent, as aerobic respiration produces more ATP per substrate oxidized. The energy produced by anaerobic respiration depends primarily on the oxidizing power of the terminal electron acceptor, with highly oxidizing compounds such as NO_3^- being nearly as efficient as aerobic

respiration, whereas fumarate respiration produces only a single ATP per succinate produced. Anaerobic respiration generates much more ATP than fermentation; however, it does require substantial investment in oxidative enzymes, respiratory complexes, and electron carriers. It is thus common in substrate-limited environments, where fermentation is not viable. Anaerobic respiration of sulfate or nitrate is very common bacterial strategy for oxidation of the organic acids and H₂ produced by fermenting bacteria.

I.D Alternatives to oxygen in anaerobes

In addition to serving as a terminal electron acceptor, O₂ plays an important role as a reactant in both catabolic and anabolic pathways. O₂ is typically activated by a reduced transition metal ion—Fe(II), Mn(II), or Cu(I)—or organic cofactor—most commonly a flavin. Within the active site of an enzyme, activated oxygen species such as peroxy radicals and Fe(IV) oxo compounds are able to abstract hydrogen atoms from unactivated substrates, thus initiating a wide range of radical chemistry. Both strict and facultative anaerobes circumvent the need for O₂ through the use of radical cofactors, often in pathways that are unrelated to their aerobic counterparts. An example that is relevant to the work in this thesis is the parallel pathways for toluene degradation in aerobic (Figure I.3A) and anaerobic (Figure I.3B) organisms (15). In aerobes such as *Pseudomonas putida* and facultative anaerobes like *Thauera aromatica*, O₂ is used at multiple steps to directly oxidize toluene, producing a substantially oxidized, linear compound that can be broken down to acetyl-CoA (16). In contrast, under anaerobic conditions toluene is inert to most cellular oxidants and thus must be functionalized to permit degradation. Benzylsuccinate synthase (BSS) catalyzes addition of toluene to fumarate, which generates functional handles in the form of the carboxylate groups. A modified β oxidation scheme generates benzoyl-CoA, which is a central intermediate in anaerobic aromatic compound degradation. Instead of breaking the aromatic ring through oxidation, benzoyl-CoA is reduced by an ATP-powered reductase. Subsequent steps hydrate, oxidize, and hydrolyze the cyclohexadiene intermediate to produce, as in the aerobic pathway, a linear compound that is easily oxidized to acetyl-CoA by β oxidation. Organisms that contain the toluene degradation pathway typically also directly utilize benzoate and have separate operons and regulation of gene expression for the two substrates. The pathway initiated by BSS is thus a branch of benzoate metabolism. Other branches—for example from *p*-cymene, xylenes, cresols, ethylbenzene, or

phenol—also lead to benzoyl-CoA, either through succinate derivatives or through other pathways. Most organisms appear to be specific for the particular aromatic compound they degrade, having a limited set of branch pathways, but almost all organisms that go through a benzoyl-CoA intermediate can utilize benzoate (17-19).

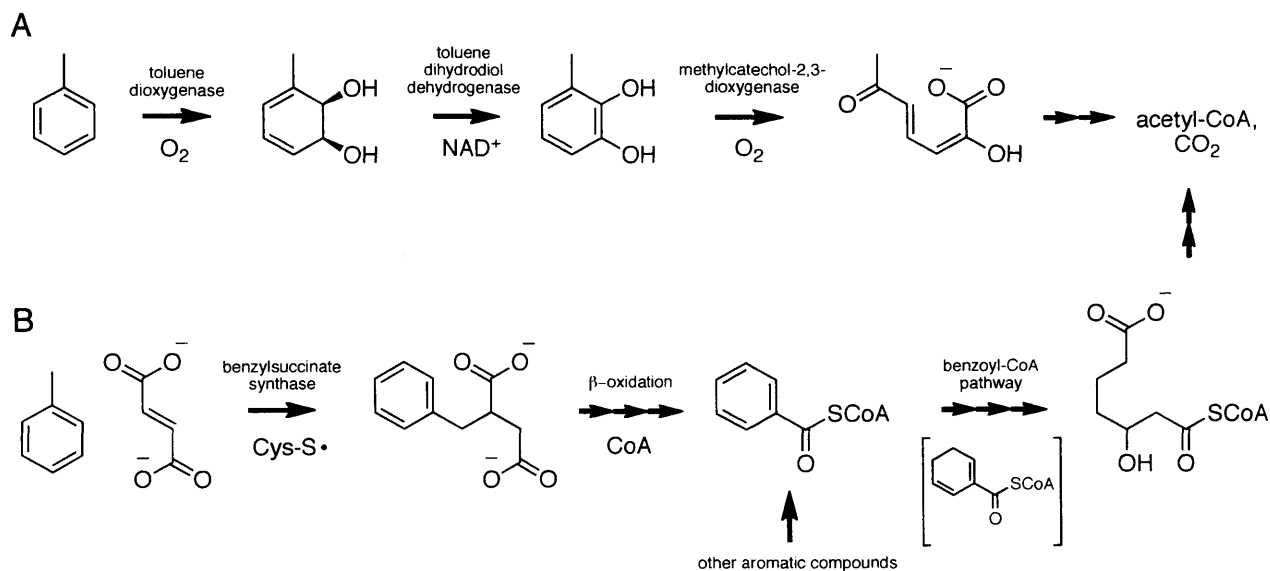


Figure I.3. Bacteria require multiple strategies to initiate metabolism of toluene in different environments. (A) Aerobic toluene degradation requires two direct oxidation steps and results in the formation of products that can be degraded to acetyl-CoA. **(B)** In the absence of O_2 , toluene degradation is initiated by conjugation to fumarate. The resulting benzy succinate enters benzoate metabolism through β oxidation to benzoyl-CoA. In contrast to the aerobic system, dearomatization of the aryl ring is accomplished through reduction rather than oxidation.

Beyond its role in catabolic pathways, bacteria also use O_2 as a reactant in biosynthetic reactions of cofactors, e.g. heme, thiamin, quinones (20). In the absence of O_2 , facultative aerobes typically switch to backup pathways that are O_2 -independent, often through enzymes that are themselves O_2 -sensitive. One example germane to this thesis is that of ribonucleotide reductase (RNR), a central biosynthetic enzyme responsible for de novo dNTP production for DNA synthesis and repair. RNRs are ancient enzymes, and are proposed to have facilitated a transition from RNA genomes to DNA (21, 22). RNRs are split into three classes based on their means of generating a catalytically essential thiyl radical within the enzyme active site (Figure I.4A). Class I enzymes use O_2 and binuclear metal center to generate a stable tyrosyl radical/oxidized metal ion cofactor within a radical storage protein, class II enzymes use a cobalamin (Cbl) derivative, adenosylcobalamin (AdoCbl), to generate a transient 5'-

deoxyadenosyl (5'-dA) radical, and class III enzymes carry a stable, backbone glycy radical, which is generated by an *S*-adenosylmethionine (AdoMet) radical enzyme. Each of these radical cofactors is proposed to abstract a hydrogen atom from an active site cysteine during turnover to form the catalytic thiyl radical, which in turn initiates chemistry on a ribonucleotide substrate. Because class I enzymes use O₂ as a substrate in their activation reaction, they are restricted to obligate and facultative aerobes and microaerophiles. Virtually all eukaryotes also use class I enzymes, but it is thought that an early eukaryotic ancestor acquired this enzyme from a bacterial source (23). Class II enzymes are much more versatile, as they neither require nor are sensitive to oxygen, and are widely distributed among bacteria; however, they require a source of AdoCbl, which many bacteria are unable to make *de novo* and is rare in many environments. Class III RNRs are extremely O₂ sensitive, but for many strict anaerobes and facultative enteric bacteria, they are an excellent option when neither Cbl nor O₂ is available. Bacteria often carry genes for multiple RNRs (24, 25) that are expressed under specific conditions: anaerobic/aerobic (26), AdoCbl-replete/-starved (27) and iron-replete/-starved (28) transitions in RNR expression have been documented. Because RNRs have developed three different mechanisms to generate a radical, each of which has unique requirements and limitations, RNR evolution and acquisition through horizontal gene transfer has enabled organisms to adapt to new environments with which they were previously incompatible (24).

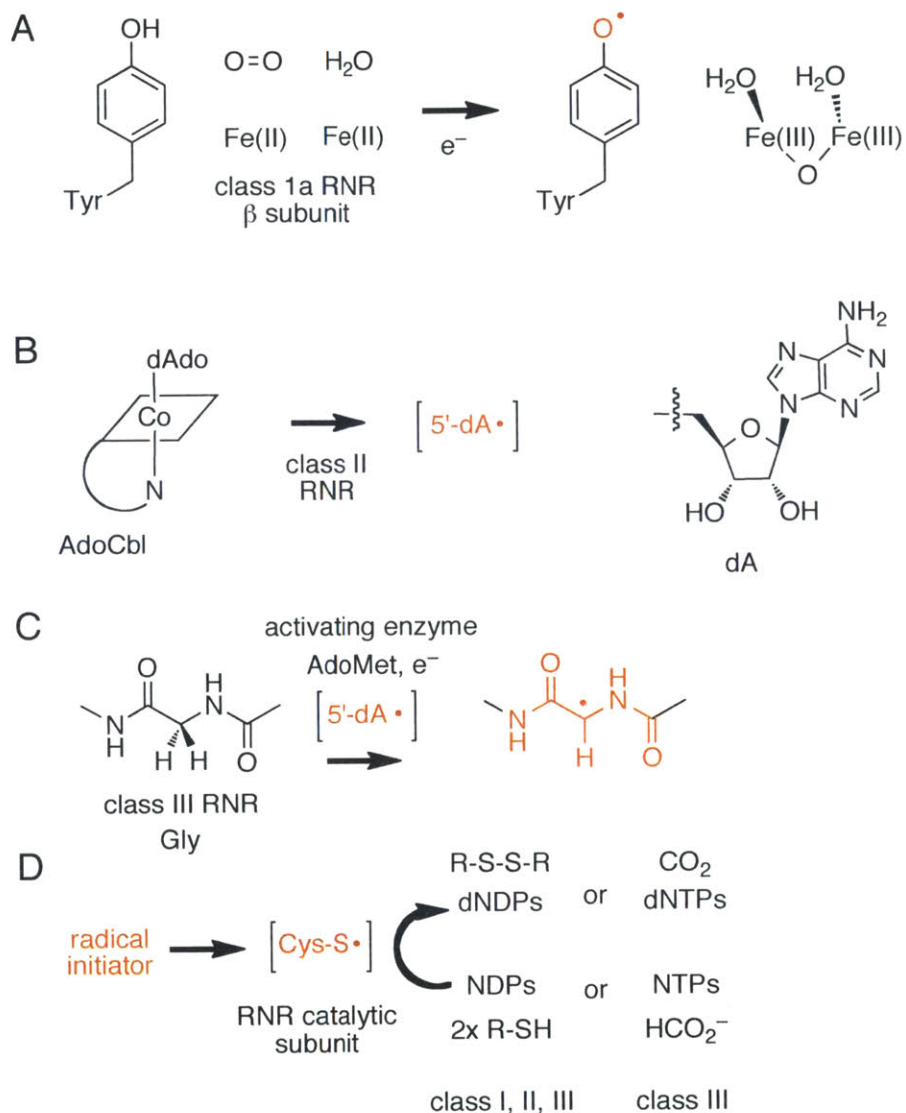


Figure I.4. Ribonucleotide reductases (RNRs) require different radical cofactors depending on environment. (A) Class I RNRs require O_2 and reduced metal ions [class Ia, Fe(II); class Ib, Mn(II); class Ic, Mn(II)-Fe(II), Tyr=Phe] for initial radical generation. The class Ia cofactor is shown. **(B)** Class II RNRs are O_2 independent, but instead require a complex and expensive cofactor, adenosylcobalamin (AdoCbl). Radical generation occurs by homolysis of the Co–C bond to yield a 5'-deoxyadenosyl (5'-dA) radical. **(C)** Class III RNRs are activated by a partner activating enzyme (see Section I.G) with a universally present and relatively inexpensive cofactor, S-adenosylmethionine (AdoMet). The activation process generates a transient 5'-dA radical, which generates a stable glycyl radical within the protein that is the radical source for catalysis. **(D)** Once the stable radical initiator has been formed (class I and III) or bound to the enzyme (class II), a transient thiyl radical is generated which initiates catalysis on NDP or NTP substrates (described in detail in Chapter V).

I.E Glycyl radical enzymes are key players in anaerobic metabolism

The primary focus of this thesis is a family of enzymes that utilize an organic radical cofactor to initiate chemical reactions otherwise impossible to accomplish with non-radical catalysts. The radical cofactor is a **glycyl radical**, formed within the same polypeptide chain as the rest of the enzyme, and this family is collectively known as **glycyl radical enzymes**, or **GREs**. The glycyl radical is ideal as a radical storage device as it is substantially less reactive than a typical organic radical. This stabilization is thought to be due to a captodative effect in which the adjacent electron-donating nitrogen and electron-accepting carbonyl synergistically lower the reduction potential of the glycyl radical (29).

The chemistry catalyzed by GREs is diverse, including C–C, C–N, and C–O bond cleaving reactions, but all GREs initiate chemistry through a common intermediate: a **thiyl radical** (Figure I.5A). An active site cysteine conserved in all GREs is thought to be transiently oxidized by the glycyl radical, a moderately uphill reaction. The transient thiyl radical, which has not yet been directly observed in a GRE system, then serves as the active oxidant, in most systems by hydrogen atom abstraction from a substrate (Figure I.5B). Once a **substrate-based radical** has been formed, residues within the GRE active site catalyze a transformation specific to the enzyme to form a **product-based radical**. The active site cysteine is then reoxidized to the thiyl radical and, in turn, oxidizes the glycine to the glycyl radical.

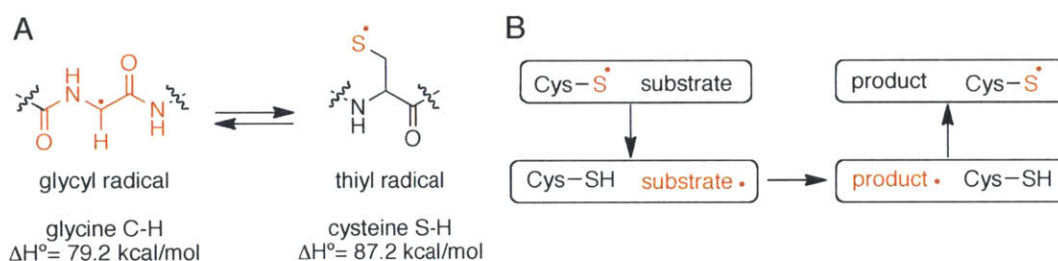


Figure I.5. Glycyl radical enzymes (GREs) utilize a series of radical intermediates to accomplish challenging chemistry. (A) A backbone glycyl radical, stabilized by delocalization and captodative effects from the two adjacent amide bonds, is inserted post-translationally in these enzymes (see Figure I.9). This glycyl radical transiently generates a thiyl radical on a cysteine in the active site. **(B)** GREs initiate chemistry by generating a transient substrate based radical; they contain this radical and encourage conversion to a product-based radical; the product-based radical is reduced by the enzyme to regenerate the thiyl radical. The thiyl radical has never been observed in a GRE system, but is inferred to exist based on structures, deuterium exchange experiments, and the reactions of suicide inhibitors (30-32). Bond dissociation energies were estimated by ab initio quantum chemistry calculations (33, 34).

Most GREs characterized thus far play a crucial role in initiating new branches of anaerobic primary metabolism. The use of a radical cofactor permits a host of unusual chemistry that is otherwise inaccessible, and many enzymes use this expanded reactivity to catalyze particularly difficult chemistry. The reactions include the rearrangement or activation of an otherwise recalcitrant molecule to permit degradation through established pathways; radical *umpolung* reactions, in which the polarity of a molecule is reversed by $1 e^-$ oxidation or reduction to permit cleavage; and radical-mediated ketone reduction with at least two different reductants.

As of 2014, six GRE reactions have been functionally verified and characterized:

Pyruvate formate-lyase (PFL, EC 2.3.1.54) initially catalyzes reversible cleavage of pyruvate to formate with acetylation of an active site cysteine (36, 37) (Figure 1.6A). Subsequently coenzyme A (CoA) can accept the acetyl group to form acetyl-CoA. PFL is central in bacterial sugar fermentation, accepting pyruvate from glycolysis. However, PFL is also the initiator for new metabolic pathways as it produces acetyl-CoA, a central metabolic intermediate that can be used in biosynthesis or consumed to produce ATP directly by substrate-level phosphorylation. PFL is crucial in mixed-acid fermentations by enteric bacteria, and the *E. coli* enzyme is the prototype. Formate produced by PFL is oxidized by the formate dehydrogenase/hydrogenase complex in acidic medium as a means of generating a proton gradient (38) but can be used for biosynthetic reactions, including ribonucleotide reduction and de novo purine synthesis (39).

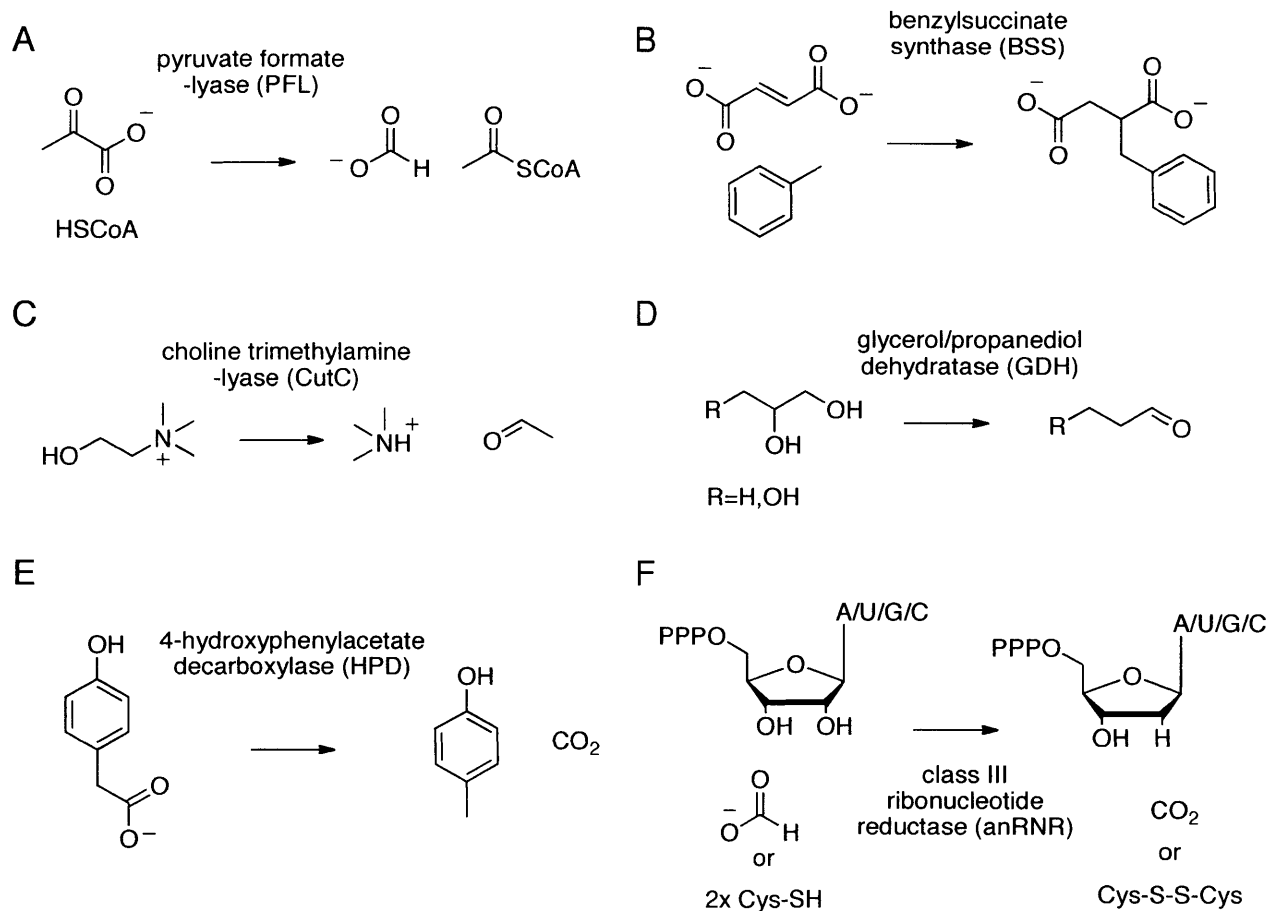


Figure I.6. GREs catalyze diverse reactions. (A, B, E) C–C lyases cleave or generate bonds between carbons. (C, D) C–O and C–N lyases catalyze 1,2-elimination reactions, forming an aldehyde. (F) Ribonucleotide reductases are unique in that they catalyze a dehydration reaction coupled to reduction of the resulting ketone.

Benzylsuccinate synthase (BSS, EC 4.1.99.11) catalyzes the addition of toluene to fumarate, generating benzylsuccinate (40) (Figure I.6B). BSS performs the first reaction in a multi-step pathway that feeds into benzoate metabolism through enzymes likely co-opted from β oxidation (41). Related enzymes are thought to process 2-methylnaphthalene (42), various cresols (43, 44), and even *n*-alkanes (45). We propose these enzymes be collectively called **X-succinate synthases (XSSs)**. The *tut* (toluene utilization) operon of *T. aromatica* contains two additional small genes that are transcribed and translated alongside the large catalytic BSS subunit. These small proteins bind [4Fe-4S] clusters and form a 1:1:1 complex with the catalytic subunit. The *tut* operon contains at least two more conserved genes of uncertain function that have been shown to be essential for BSS function in vivo. All of these additional genes appear to

be broadly conserved in XSS gene clusters. Although some XSSs have been characterized in crude cell extracts (primarily BSS), no activity has yet been observed in a fully reconstituted in vitro system, despite the successful expression and purification of BSS from *E. coli* (46). XSSs are known to function primarily in nitrate-reducing γ -Proteobacteria, and sulfate-reducing δ -Proteobacteria, although they have also been found in anaerobic phototrophic bacteria (47), metal-ion-reducing *Geobacter metalireducens* (48) and in syntrophic communities of methanogenic archaea and sulfate-reducing bacteria (49). XSS enzymes have thus permitted incorporation of hydrocarbons into many different kind of bacterial metabolism.

Choline TMA-lyase (CutC, EC 4.3.99.4) cleaves choline to trimethylamine (TMA) and acetyldehyde as the first step in anaerobic choline degradation (50) (Figure I.6C). The *cut* operon (choline utilization) contains alcohol and aldehyde dehydrogenases, capable of acetaldehyde disproportionation to acetyl-CoA and ethanol (51). Interestingly, these genes in the *cut* operon are homologous to those used for AdoCbl-dependent ethanolamine and propanediol degradation in the *eut* and *pdu* operons. The *cut* operon also contains a set of structural proteins that are thought to assemble into cellular microcompartments, possibly decreasing the toxicity of the acetaldehyde this enzyme produces. CutC contains an N-terminal conserved sequence that may be involved in its packaging into these compartments. It appears that CutC supports organisms through fermentation or sulfate respiration, and is found commonly in enteric proteobacteria and clostridia. The only in vitro characterized enzyme thus far comes from *Desulfovibrio alaskensis* (52).

Glycerol dehydratase (GDH, EC 4.2.1.30) cleaves glycerol to water and 3-hydroxypropanal (3-HPA), which, in *Clostridium butyricum*, is then reduced by NADH to generate propane-1,3-diol as a byproduct of fermentation (53) (Figure I.6D). GDH is encoded in the *dha* operon (dihydroxyacetone inducible), which includes an alcohol dehydrogenase. Most GDH enzymes can also dehydrate propane-1,2-diol to propanal, and some may perform this reaction preferentially during degradation of certain sugars in analogy to AdoCbl-dependent diol dehydratase (54). GDH primarily functions in fermentation by clostridia and enteric γ -proteobacteria, and the only enzyme characterized thus far is from *Clostridium butyricum* (55).

4-hydroxyphenylacetate decarboxylase (HPD, EC 4.1.1.83) cleaves the tyrosine metabolite 4-hydroxyphenylacetate to CO₂ and *p*-cresol (56, 57) (Figure I.6E). Although this reaction does not contribute to central metabolism, *p*-cresol is believed to promote virulence of

Clostridium difficile through inhibition of the growth of other bacteria that would compete with it for nutrients (58). Similar to BSS, HPD is composed of a large, catalytic subunit and a small, multi-[4Fe-4S]-cluster-binding subunit of unknown function. Additionally, there is limited data suggesting that HPD forms a heterooctameric structure dependent upon the small subunit and is subject to regulation by phosphorylation (57), but these claims are preliminary. Unlike BSS, however, the *hpd* operon is not reported to contain any additional conserved genes that could contribute to the function of the enzyme or its regulation. HPD is found primarily in pathogenic clostridia, and the enzymes from *Clostridium difficile* (59), *Clostridium scatologenes* (60, 61) and *Tannerella forsythensis* (61) have been characterized.

Class III or anaerobic ribonucleotide reductase (class III RNR, EC 1.17.4.2) reduces ribonucleoside triphosphates (NTPs) to 2'-deoxynucleotides (dNTPs) with concomitant oxidation of either formate (62) or an active site thiol pair (35) (Figure I.6F). Class III RNR is not involved directly in energy metabolism; however, it is the sole source of substrates for DNA synthesis and repair, is an important consumer of reducing equivalents, and is linked to central metabolism through its use of formate or reduced thiols. Class III RNR is distributed throughout bacterial phyla and are also found in some archaea, primarily the strictly anaerobic methanogens. The enzymes from *Escherichia coli*, *Lactococcus lactis*, *Thermotoga maritima*, and T4 bacteriophage have been characterized.

I.F Core features of glycy radical enzyme structure and catalysis

GREs, like adenosylcobalamin (AdoCbl) and *S*-adenosylmethionine (AdoMet) radical enzymes, utilize a **barrel architecture** to encapsulate the active site of the enzyme (Figure I.7A-E). This architecture has been proposed to protect the radical-generating cofactors as the active site is typically completely shielded from solvent (63). AdoCbl radical enzymes, with the exception of class II RNR, are $(\beta/\alpha)_8$ complete triosephosphate isomerase (TIM) barrels with a AdoCbl-binding domain that caps the catalytic domain. AdoMet radical enzymes are based on the same TIM barrel core, but are often abbreviated to $(\beta/\alpha)_6$ barrels, possibly to engage large substrates such as proteins and nucleic acids or to accommodate accessory domains that expand the reactivity of the enzyme (64). AdoMet and AdoCbl barrels appear to be homologous, especially in light of the discovery of ThiC, an AdoMet radical enzyme with a domain architecture very similar to some AdoCbl enzymes (65).

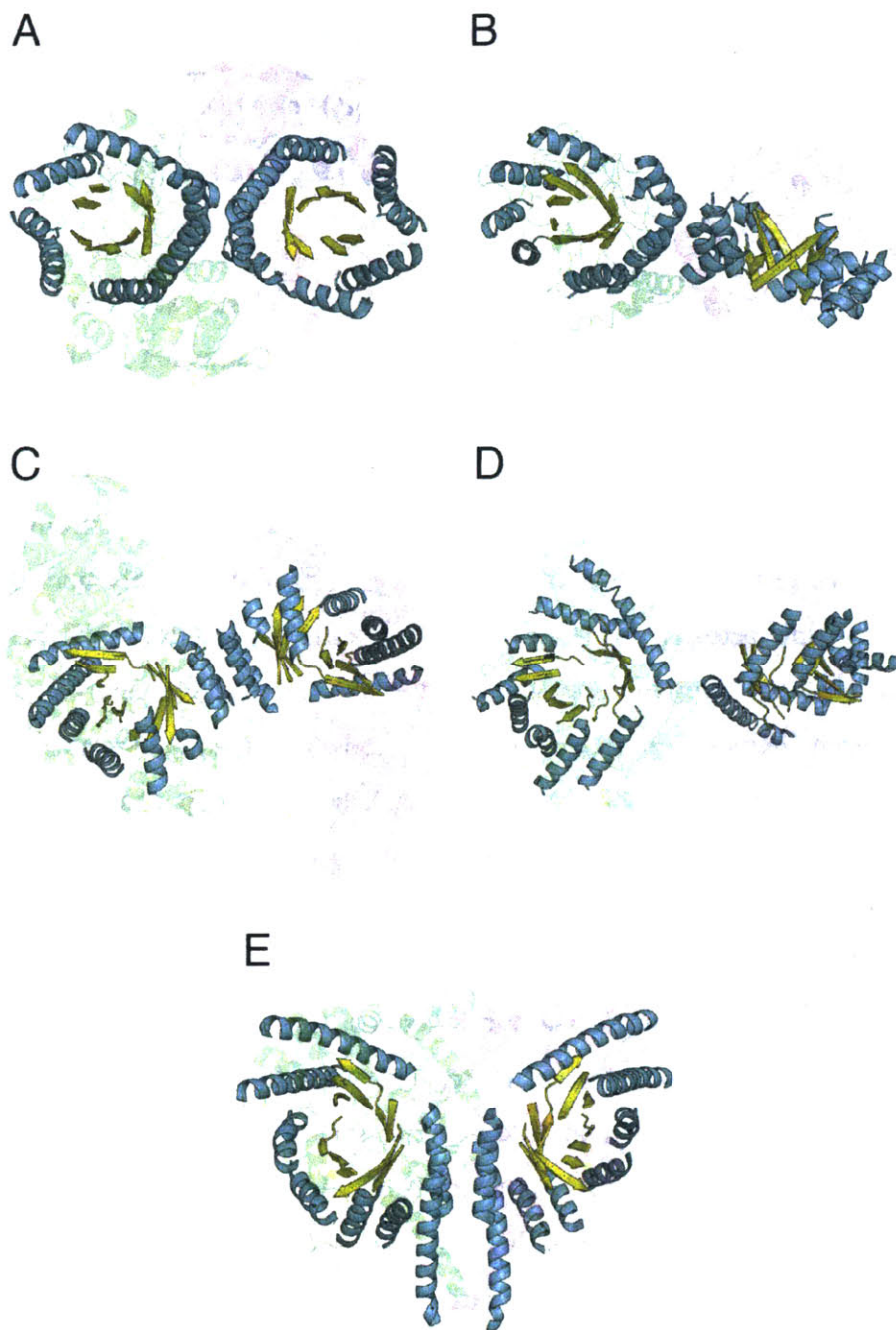


Figure I.7. Examples of radical enzyme architecture based upon a β/α -barrel fold. A-E) Core strands (yellow) and helices (cyan) are distinguished from non-core features of each protomer (transparent green or magenta). **(A)** AdoCbl radical enzymes use a complete, 8-stranded TIM barrel with a dimeric core structure. AdoCbl-dependent ethanolamine ammonia-lyase is shown [PDB ID 3AO0 (66)]. An additional protein chain binds AdoCbl. **(B)** AdoMet radical enzymes use an incomplete TIM barrel with 5, 6, or 7 strands, or a complete barrel with 8 strands. Biotin synthase is shown [PDB ID 1R30 (67)]. Some AdoMet radical enzymes adopt a dimeric architecture, but not all use the same dimer interface, and many are monomeric. **(C)**

Class I and II RNRs use a 10-stranded barrel with a dimeric interface generated by two of the core helices. *E. coli* class Ia is shown [PDB ID 1RLR (68)]. **(D)** GREs use a 10-stranded barrel with a dimer interface created by one core helix and three external helices. *E. coli* pyruvate formate-lyase is shown [PDB ID 2PFL (69)]. **(E)** Class III RNRs use a 10-stranded barrel with a dimer interface created by core helices, but with a topology resembling the GREs, rather than the class I and II RNRs. The formate-dependent enzyme from T4 bacteriophage is shown [PDB ID 1H78 (70)].

In contrast to the all-parallel fold of the TIM barrels, class I and class II RNRs and all known GREs form a superfamily of enzymes that contain two 5-stranded parallel β/α half-barrels arranged head-to-tail as a complete, 10-stranded β barrel (Figure I.7C-E). The α helices in RNRs and GREs have undergone extensive rearrangement, and two core helices are missing entirely from the β/α architecture, whereas numerous non-core helices have been introduced. The order of evolutionary divergence of the different RNRs and GREs are currently unclear. There is no clear evolutionary link between the 8-stranded TIM barrels and GRE/RNR barrels, but it is tempting to speculate that they may be related given the common chemistry that they catalyze and the link between GREs and their partner AdoMet radical activating enzymes (AEs).

At the center of the GRE barrel are two loops, the **Gly loop** and the **Cys loop**, which contain the universally conserved glycine and cysteine residues necessary to initiate catalysis in every GRE (Figure I.8A). The Cys loop bridges the two half-barrels, following $\alpha 5$ and preceding $\beta 6$, and inserts directly in the middle of the barrel, filling the center of the barrel from one side (Figure I.8B). The Gly loop is present at the innermost tip of a structural element at the very C terminus of the protein: the **glycyl radical domain**. This domain comprises a short β hairpin flanked by two helical segments that fold around a hydrophobic core, providing a stable, independent domain. In all class III RNRs, there is an additional Zn-ribbon motif that replaces the first helix, but the overall topology of the domain is similar. The Gly loop is inserted into the barrel on the opposite face as the Cys loop, resulting in the two loops contacting each other in the middle to enable direct radical transfer between glycine and Cys. In most GREs, specific residues that contribute to catalysis are found on the Cys loop, which also typically helps form the binding pocket for substrate. In PFL this includes a second cysteine immediately preceding the first that is essential for turnover and thought to be involved in accepting the acetyl group from pyruvate (Figure I.8C). In dehydratase enzymes (CutC, GDH, and related) a glutamate residue present two residues after the cysteine appears to be an essential general base (Figure I.8D). The interior face of the β strands contributes to the **substrate binding pocket** as well as

stabilizing the Gly and Cys loops (Figure I.8B). Strands $\beta 1$, $\beta 2$, $\beta 6$, and $\beta 7$ provide residues important to substrate binding and catalysis in most GREs and are highly variable both in sequence and in structure. $\beta 3$ -5 and $\beta 9$ -10 are found on the underside of the two active site loops and are the least variable across the family.

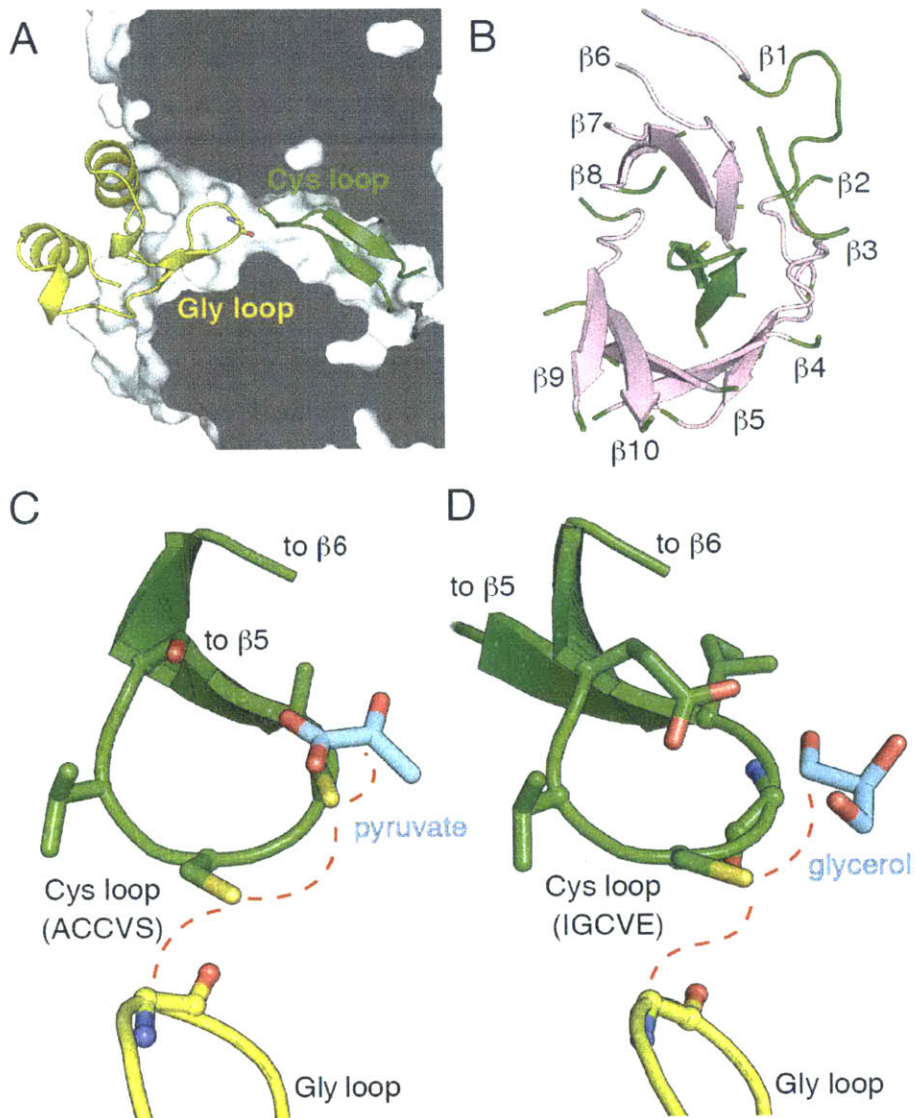


Figure I.8. GREs contain two deeply buried loops that carry residues essential for glycy radical storage and catalysis. (A) The Gly and Cys loops contain the glycine used for radical storage and the cysteine used for radical initiation. The Gly loop is found on the glycy radical domain at the C-terminus of the protein. **(B)** The core GRE barrel is composed of 10 β strands (pink) that encircle the Cys loop. The active site cavity is entirely contained between the Cys loop and the top of the barrel, with residues from $\beta 1$, $\beta 2$, $\beta 6$, and $\beta 7$ dominating the upper surface. Several of these strands are highly distorted in regions (green). The bottom of the

barrel stabilizes the Cys and Gly loops and is typically highly conserved between GREs. **(C)** The active site of pyruvate formate-lyase and **(D)** glycerol dehydratase are shown with bound substrate (cyan). Hydrogen-atom transfer (red lines) is thought to occur between the glycy radical Ca, the cysteine thiol, and the substrate as diagramed in Figure I.5.

Structurally, the GRE family is broadly split between **PFL-like** and **class III RNR enzymes**. These two groups share the common GRE fold, described above, but also contain distinct features. All GREs thus far characterized undergo **dimerization** through a dimer interface that varies among subfamilies. There is some indication that a tetrameric unit may be common in GDH and its relatives as many of these enzymes adopt a tetrameric structure in crystal lattices (Chapter IV), and some members have been characterized as tetramers in solution. PFL/GDH-like enzymes contain two long, antiparallel helices inserted between $\beta 1$ and $\alpha 1$ that form a dimer interface $\alpha 6$, which is extended beyond the typical helix length. (Figure I.7D). Class III RNRs contain a single additional helix inserted between $\beta 2$ and $\alpha 2$, and incorporate $\alpha 1$ and $\alpha 2$, aligned in parallel, to form part of a dimerization interface that includes allosteric nucleotide binding sites (Figure I.7E). A long helix that precedes the barrel core completes the interface with numerous cross-dimer contacts. The class III RNR interface does not appear to be related in either sequence or structure to that observed in other GREs, suggesting an independent origin for the two enzyme groups from a progenitor GRE, perhaps from a monomeric ancestor. The class I/II RNRs also have a completely different interface from the class III RNRs (Figure I.7C, I.7E). In class I/II enzymes, $\alpha 1$ and $\alpha 2$ alone generate the dimer interface, but are arranged antiparallel instead of parallel. In both cases, specificity regulation is mediated by nucleotide binding near the dimer interface and follows similar rules. It is proposed that the mechanism for this regulation is conserved; however, the structural elements responsible for communication between the specificity and active sites are quite different (70). A notable exception to the dimeric structure of RNRs is the class II enzyme from *Lactobacillus leichmanii*, which is a monomer but contains a short insertion that recapitulates the dimeric interface observed in other class II and class I enzymes.

I.G Glycyl radical enzyme activation by activating enzymes

In order to be catalytically active, GREs must undergo an initial activation step in which the glycy radical is installed by an activating enzyme (AE). This step is specific, as each GRE has a partner AE that is typically encoded adjacent to it in the genome. GRE-AEs are members

of the AdoMet radical enzyme superfamily, all of which are thought to generate a transient 5'-dA radical from AdoMet and a reduced [4Fe-4S] cluster (Figure I.9). The 5'-dA radical directly abstracts a hydrogen atom from an unactivated carbon on a substrate, forming a substrate-based radical. For AEs, the substrate is the C α of the active site glycine of the GRE, and no additional chemistry involving the AE occurs once this stable radical is formed.

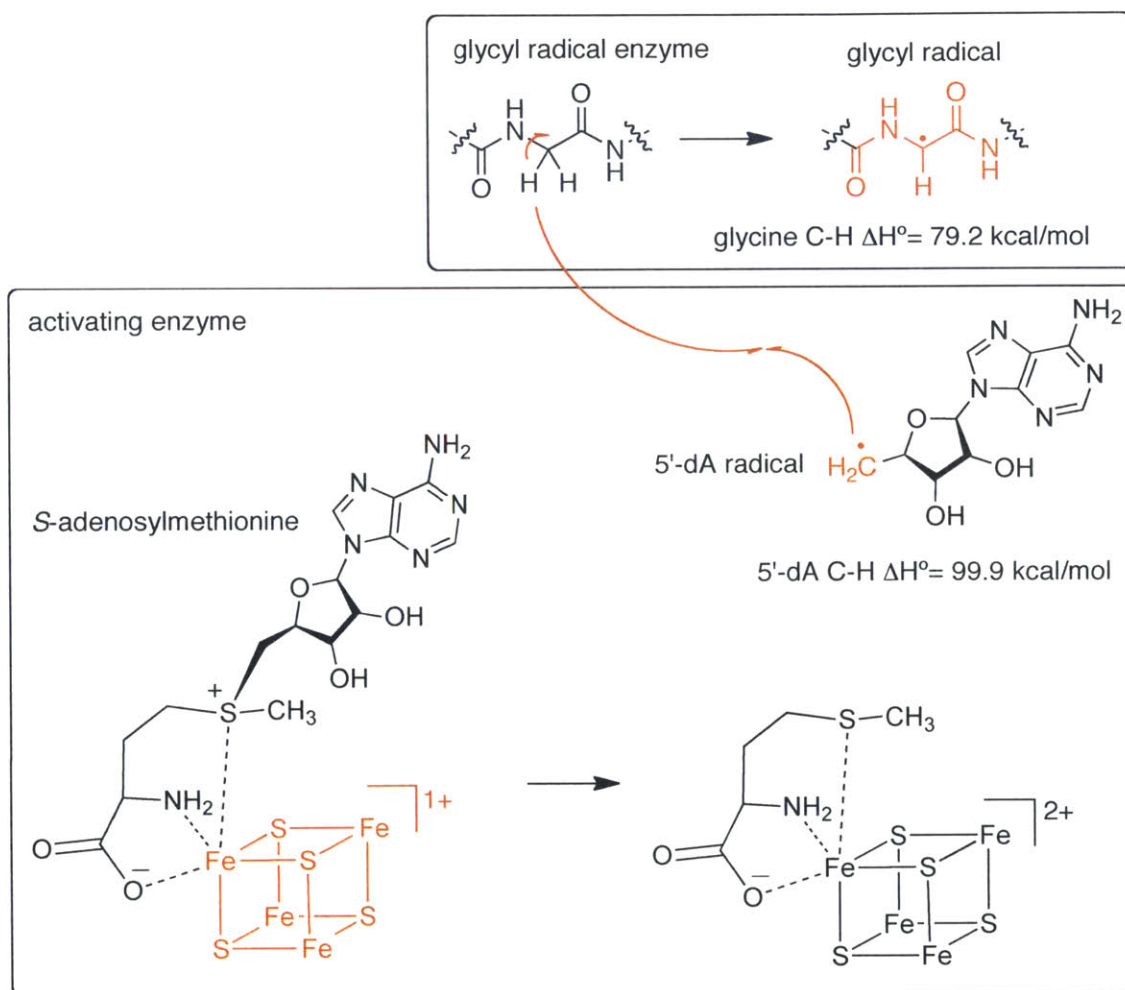


Figure I.9. Activation of a GRE by its partner activating enzyme. Radicals and radical transfer steps are indicated (red lines). The [4Fe-4S]¹⁺ [$E^\circ = -600$ mV, bound to AdoMet (71)] reductively cleaves AdoMet [$E^\circ = -990$ mV, bound to enzyme (72)] to generate a 5'-dA radical that directly abstracts a hydrogen atom from the GRE glycine. The C-H bond dissociation energies of 5'-dA and glycine have been estimated by ab initio molecular orbital calculations (34, 73).

Like all members of the AdoMet radical enzyme superfamily, the basic fold of AEs is based on a TIM (triose phosphate isomerase) barrel (Figure I.10A), which contains eight β/α

units arranged in an antiparallel closed barrel. All AEs, and many other AdoMet radical enzymes, have a partial barrel containing anywhere from five to seven strands. PFL-AE is the prototype for AEs, and the structure (Figure I.10A-D) will guide the discussion below. In the active site of nearly all AdoMet radical enzymes, three cysteine residues are present within a conserved **CX₃CX₂C motif** (Figure I.10B), which is found in a loop between the first β strand (β 1) and α helix (α 1) of the core AdoMet radical barrel. These cysteines coordinate a [4Fe-4S] cluster leaving an open coordination site facing the interior of the barrel. Electron paramagnetic resonance (EPR) and electron nuclear double resonance spectroscopies first suggested that AdoMet binds to the [4Fe-4S] cluster directly in PFL-AE (74). The structure of biotin synthase bound to AdoMet confirmed that the exposed Fe binds directly to the amine and carboxylate of AdoMet, forming a square-pyramidal coordination environment around the Fe atom (67). The open site at the bottom of the pyramid is occupied by the sulfonium of AdoMet. Upon reduction of the [4Fe-4S] cluster from the 2+ to 1+ oxidation state by an external electron donor, an electron is injected into the C-S σ^* orbital of AdoMet through an inner-sphere electron transfer mechanism (75). The reduction of the sulfonium results in homolytic bond cleavage of the C-5'-S bond and formation of the 5'-dA radical and methionine, which is now coordinated directly to the unique Fe through the new thioether lone pair. The 5'-dA radical is an extremely reactive, primary radical, capable of removing a hydrogen atom from an unactivated carbon. In some cases, AdoMet is regenerated by the enzyme after each turnover as in AdoCbl enzymes; however, in most AdoMet radical enzymes the 5'-dA radical cannot be regenerated as the reverse hydrogen-atom transfer is too far uphill. AdoMet is thus used as a co-substrate to generate an organic radical that is typically quenched by $1e^-$ oxidation to yield product (76). For GRE-AEs, the substrate is a glycine residue on the GRE, and quenching of the radical appears to be suppressed. In contrast to the 5'-dA radical, the glycy radical is easily regenerated every round of turnover within GREs, making it suitable for central metabolic processes which require high turnover numbers and flux in order to power cellular energy conservation. The mechanism by which AdoMet and AdoCbl are regenerated in enzymes that use them catalytically is still mysterious but is likely linked to the ability of product radical to regenerate the 5'-dA radical or coupling of C-H and C-Co or C-S bond forming and breaking reactions.

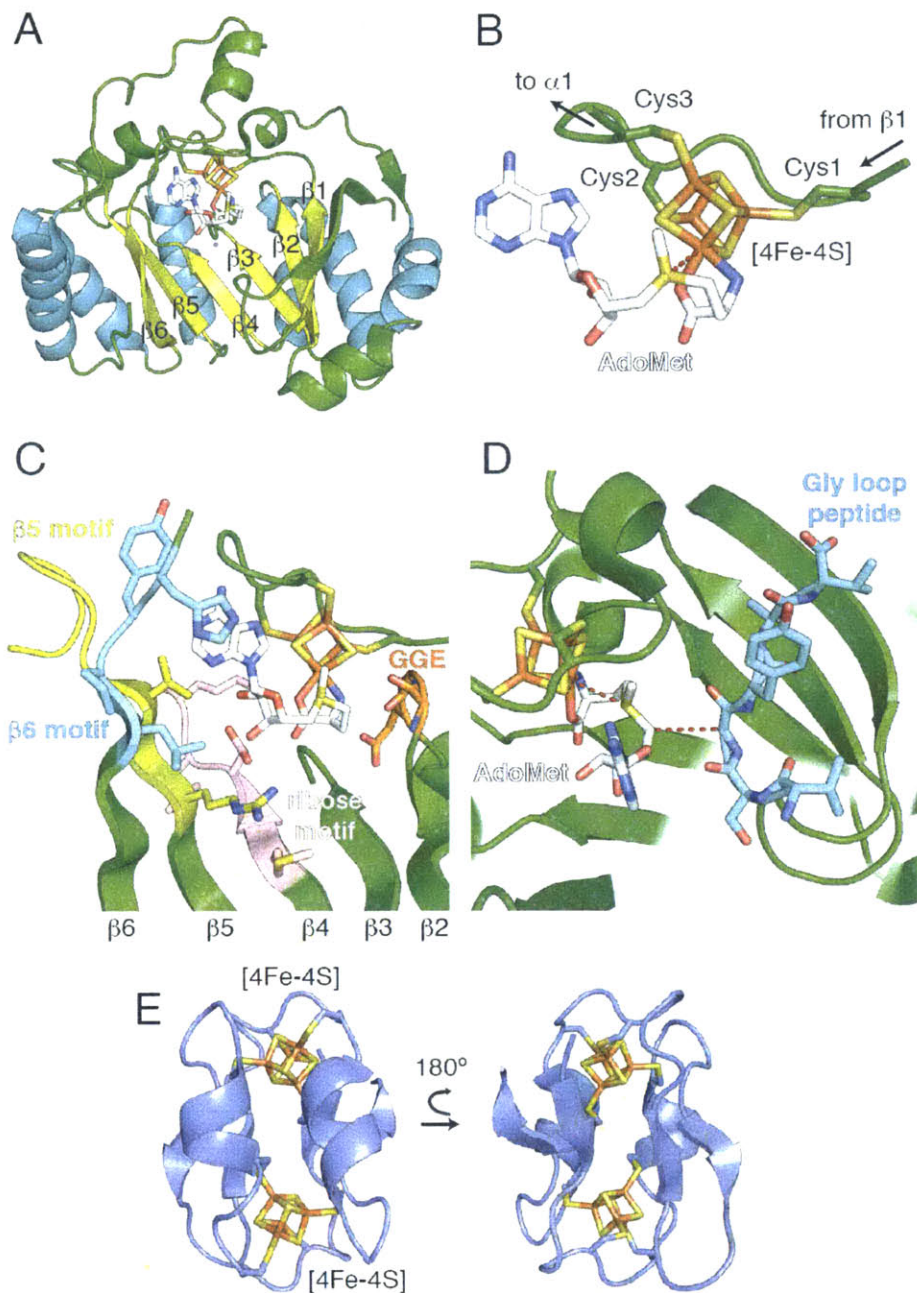


Figure 1.10. AdoMet radical activating enzymes (AEs) generate the active site glycyl radical in GREs. A-D) Pyruvate formate-lyase (PFL) AE [PDB ID 3CB8 (77)] is the only AE for which a structure with bound cluster and AdoMet are available. The core helices (cyan) and strands (yellow) are highlighted **A)** The PFL-AE barrel contains 6 core strands arranged in a partial barrel. **(B)** The [4Fe-4S] cluster is coordinated by three cysteines within a conserved loop between $\beta 1$ and $\alpha 1$. AdoMet (white) coordinates the unique, exposed iron through its amino and carboxylate groups. The sulfonium is positioned above the cluster in the AdoMet-bound structures. **(C)** Four motifs contribute to AdoMet binding, as described in the text. **(D)** A Gly loop peptide corresponding to 7 residues of the Gly loop of PFL binds in a cleft within the AE directly

adjacent to the 5'-C. (E) A partial or complete bacterial ferredoxin domain is predicted within the vast majority of GRE-AEs. A homologous ferredoxin domain from *Clostridium acidurici* is shown (PDB ID 2FDN (78)).

In addition to the CX₃CX₂C motif, there are four additional motifs (Figure I.10C) present in most, but not all, AdoMet radical enzymes that facilitate binding and cleavage of AdoMet, as first revealed by the structure of AdoMet bound to biotin synthase (64, 67, 79):

- The **GGE motif** is found at the end of β 2 and coordinates the amino nitrogen of AdoMet with a backbone amide contact. This motif, with the sequence GGE, is conserved in all AEs, but varies widely in the broader AdoMet radical superfamily.
- The **ribose motif** is found on β 4 and involved in coordination of the 2'- and 3'-hydroxyl groups of AdoMet. This motif is very well conserved in the PFL-AE-like AEs where the sequence is hhDhK (where h is a hydrophobic residue). It is present but much more variable among RNR-AEs.
- The **GXIXGXxE motif** or **β 5 motif** is so named for its sequence in biotin synthase, but in the AEs there is little consensus even within the PFL-AE-like enzymes. It begins within β 5 and continues all the way through the linker to α 5. Some RNR-AEs appear to retain this motif, yet others terminate immediately after β 5 or are not at all conserved.
- The **β 6 motif** directly coordinates the adenine ring of AdoMet through a hydrogen bond from a backbone amide and also surrounds the adenine with hydrophobic residues. This structural element occurs in some form in all AdoMet radical enzymes characterized to date, but most RNR-AEs appear to be completely missing β 6, so binding of the adenine moiety of AdoMet is likely to be somewhat different in these enzymes. One exception is a group of RNR-AEs from organisms including *Pseudomonas aeruginosa*; these enzymes contain residues corresponding to α 5, β 6, and α 6, but the β 5 and β 6 motifs are not at all conserved, suggesting that the mode of AdoMet binding may vary even within RNR-AEs.

In addition to variations in the core motifs, most characterized GREs other than class III RNR and PFL (GDH, BSS, CutC, and HPD) are activated by an AE containing a ferredoxin-dicluster domain predicted to bind two additional [4Fe-4S] clusters. These domains are ubiquitous in bacterial and archaeal redox enzymes and typically mediate electron transfer within proteins. This domain consists of an α - β sandwich fold with two CX₂CX₂CX₃C cluster-binding motifs found distributed in two α helices (Figure I.10E). The motifs are separated by 15-20

residues that form an antiparallel β sheet that shields the clusters from solvent access. Each cluster motif contributes three cysteines to one cluster (CX_2CX_2C) and the final cysteine to the other cluster. The ferredoxin domain is inserted immediately following the CX_3CX_2C motif that coordinates the AdoMet radical [4Fe-4S] cluster, suggesting that it could play a role in shuttling electrons from an electron donor to the AdoMet radical cluster. The utility of this expensive additional domain is unclear as most AE and most AdoMet radical enzymes function without any such electron shuttle. Other AdoMet radical enzymes are known to have alterations in and around the AdoMet radical cluster which, in at least one case, are believed to determine what electron donor provides the electron for reduction of the AdoMet radical cluster (80).

A puzzling feature of the GRE activation reaction is that radical generation by the AE occurs by direct hydrogen-atom transfer (81), requiring the glycine substrate and AdoMet to be within 3-4 Å, but in all structure of GREs so far determined, the glycine is buried deep within the GRE barrel, 15-20 Å from the surface of the protein. Insight into the mechanism of the activation process has primarily come from work with the best-characterized GRE/AE system, PFL/PFL-AE. The structure of PFL-AE bound to a small peptide (Figure I.10D) containing the Gly loop sequence revealed a binding site for this loop adjacent to AdoMet within the barrel of PFL-AE (77), consistent with data showing incorporation of deuterium label into 5'-dA from the same labeled peptide (81). The structures of PFL and PFL-AE bound to the peptide are thus at odds as to the conformation of the Gly loop. This incongruity led to a model for PFL activation wherein the glycy radical domain must exit the barrel of PFL to come in contact with PFL-AE and form a complex productive for activation (Figure I.11). Subsequent characterization of mixtures of PFL and PFL-AE revealed structural changes associated with complex formation, observed by CD spectroscopy; a reduction in PFL activity at low ratios of PFL to PFL-AE; and an increase in non-physiological glycy radical reduction by dithiothreitol as the amount of PFL-AE was increased (82, 83). These data suggest that PFL-AE does indeed stabilize a non-catalytic conformation of PFL where the glycy radical is more exposed to external reductants. The discovery that the binding of PFL-AE reduces the activity and accelerates loss of the glycy radical is particularly surprising given that at cellular concentrations [$\sim 20 \mu\text{M}$ PFL and $\sim 1 \mu\text{M}$ PFL-AE (36)], PFL-AE is expected to be mostly bound to PFL based on an affinity of $3 \mu\text{M}$, determined by surface plasmon resonance (84). The biggest unanswered questions in this mechanistic process—aside from the structure of the complex—are thus, 1) does the glycy

radical domain exist in an equilibrium between in and out states, or is exit catalyzed by the presence of the AE; 2) once activated, does the glycy radical domain prefer to remain in the GRE; 3) are there means of regulating this activation process through controlling access to the glycy radical domain; 4) what is the stoichiometry of activation and how is this regulated. This last question is particularly interesting as all GREs thus far isolated and activated in vitro contain, at most, 0.5 mol glycy radical per mol dimeric protein. There is, as yet, no evidence that only a single radical is generated per dimer, but both in vivo and in vitro experiments have substantiated the 1:1 stoichiometry of activated and unactivated enzymes (85-87). The mechanistic basis for this apparent half-of-the-sites activation is unclear as the glycy radical domains are separated by ~ 70 Å in GREs.

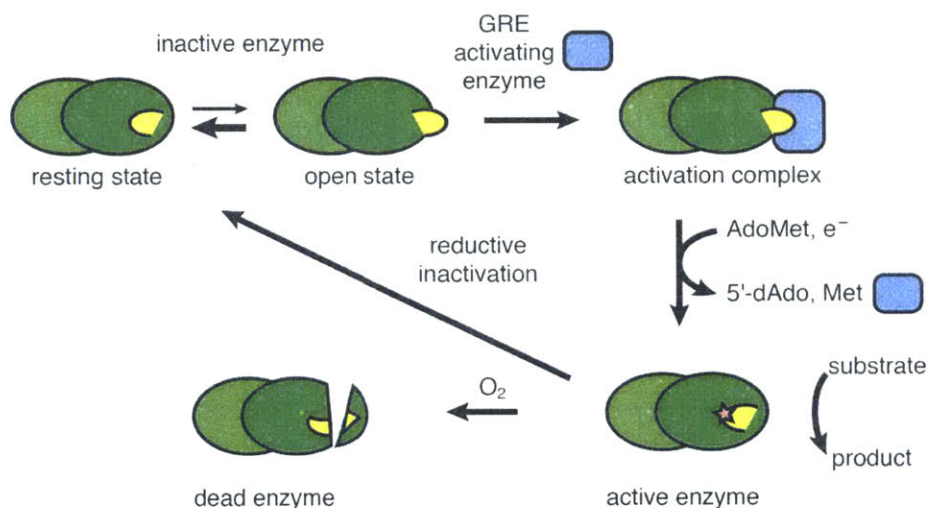


Figure I.11. GREs are post-translationally activated. Inactive enzymes are thought to exist in equilibrium between a resting (closed) and open state, in which the glycy radical domain and Gly loop are exposed. The AE binds to the open state to form an activation complex. Active enzyme is produced as described in Figure I.9. The active enzyme performs catalysis until the radical is intercepted, either by O₂ to form a dead, cleaved enzyme, or by a reductant (or loss of an oxidized, radical intermediate) to yield an inactive enzyme that can be re-activated. *This figure has been reproduced from (88).*

There are experimental difficulties in working with GREs that must be acknowledged. Firstly, generating the glycy radical cofactor requires an additional enzymatic component, the AE. Purification of AEs is quite challenging as strictly anaerobic conditions are required in order to maintain the [4Fe-4S] cluster. Additionally, from unpublished experiments in the Drennan, Marsh, and Stubbe labs, many AEs appear to be insoluble or poorly expressed in *E. coli*. AEs are

also specific for their GRE in most cases, although cross reactivity has been observed for closely related class III RNR AEs. Secondly, even when the AE is available, generation of the glycy radical can be problematic, especially for GRE active site mutants (see Chapters IV and V). Because of this issue, activity results must be interpreted with skepticism unless controls to verify the radical is present are performed. Slow radical loss has been documented in the case of HPD (89) and PFL (82, 90), although whether this process results solely from non-specific reduction by a non-physiological reductant like DTT or dithionite, or instead is a regulated, physiological process is uncertain. Regardless, the combination of inefficient glycy radical production, apparent half-of-the-sites reactivity, and slow radical loss makes obtaining kinetic parameters and designing experiments to analyze reaction mechanisms challenging. Finally, GREs lack the UV-visible spectroscopic handles of many other protein-based radical enzymes or metalloenzymes. Although the glycy radical is visible by EPR spectroscopy, on-pathway substrate- or product-based radical intermediates have never been observed in a GRE system. One partial exception is class III RNR, in which a protein-based thiosulfuranyl radical can be observed mid-way through the reaction after dehydration of the ribonucleotide (91). This intermediate is kinetically and chemically competent and only accumulates if formate is excluded from the reaction mixture. This method of observing intermediates will thus not be useful in other systems. The lack of evidence for radical intermediates in GRE systems may also be due to the inability to acquire large and highly concentrated quantities of activated GRE, a necessity for rapid freeze-quench EPR. Manual quenching is not an option in most systems due to fast ($>100\text{ s}^{-1}$) reaction rates but may be useful for experiments with mechanism-based inhibitors.

I.H Goals of the thesis

Glycy radical enzymes participate in numerous and diverse metabolic pathways in bacteria, many of which have implications in human health, the environment, and industrial biosynthesis. In addition, the study of GREs has revealed new radical chemistry that microbes use to catalyze challenging reactions, many of which may have value in industry or bioremediation. This thesis explores the structure and function of three GREs with vastly different substrates but which employ common strategies for generation and control of radical intermediates. Chapter I has introduced some of the problems and possibilities of anaerobic metabolism and described the

structural, biochemical, and microbiological characterization of a particular class of strictly anaerobic enzymes, the GREs, prior to this work.

The first experimental section, comprising Chapters II and III describes the quaternary structure and substrate binding of benzylsuccinate synthase, a GRE crucial for bioremediation of hydrocarbon-polluted environments. BSS is unusual among GREs in that along with the large, catalytic subunit and the radical generating AE, the gene cluster contains several additional proteins that are essential for its activity in vivo. Chapter II describes the structure of the BSS catalytic subunit in complex with two additional subunits that bind [4Fe-4S] clusters. One of these subunits is essential for the solubility of BSS, but the other appears to serve a regulatory role. The structure of BSS in the absence of this regulatory subunit reveals large conformational changes that may be involved in the GRE activation process. Chapter III focuses on a different aspect of BSS function: the binding of fumarate and toluene at the active site. The structure of BSS bound to these substrates reveals sequence motifs that confer selectivity and avoid unproductive binding of competing metabolites. Homology modeling of related XSS enzymes sheds light on the evolution of GREs as isolated regions of the protein have adapted to accommodate new substrates.

The second section of the thesis, comprising Chapter IV, presents structures of choline-TMA lyase, a GRE found commonly in gut commensal bacteria, which may contribute to human disease. CutC cleaves a C–N bond, which is a unique reaction so far in the GRE family. Structures of CutC with choline and TMA bound reveal the binding site and crucial residues for recognition of this substrate. Characterization of mutants of residues around the active site has revealed how loss of active site functional groups impacts binding and catalysis by this GRE.

The final section of the thesis focuses on expanding structural knowledge of class III RNR, an ancient GRE that produces substrates for DNA synthesis and repair. Despite more than 20 years of research, the second half of the reaction of anaerobic ribonucleotide reductase, the reduction of the 3'-ketonucleotide intermediate, is still mysterious, in part due to a dearth of structural information. Chapter V presents the structure of a class III RNR from *Thermotoga maritima*, revealing new details about the overall architecture of these enzymes. Accompanying biochemical data from a related enzyme suggest a subclass of anaerobic RNRs gain reducing equivalents from a redoxin instead of from formate, a new catalytic paradigm for class III RNRs.

Finally, in Chapter VI, the results of the experimental work of this thesis are summarized and placed in context of the field as a whole. Genomic databases have expanded much faster than biochemical and structural characterization, and this is especially the case for enzymes such as GREs where there are experimental difficulties in obtaining functional enzymes through heterologous expression and in working with organisms with highly specialized metabolism. Sequence clustering, analysis of gene clusters, and inferred structural homology may be able to reveal patterns that lead to effective prediction of function in some cases for uncharacterized GREs and their AEs.

I.I References

1. Halliwell B (2006) Reactive species and antioxidants. Redox biology is a fundamental theme of aerobic life. *Plant Physiol* 141(2):312-322.
2. Voet D VJ, Pratt CW (2006) *Fundamentals of Biochemistry: Life at the Molecular Level* (John Wiley & Sons) 3rd Ed.
3. Winterbourn CC (2008) Reconciling the chemistry and biology of reactive oxygen species. *Nat Chem Biol* 4(5):278-286.
4. Uden G, Becker S, Bongaerts J, Schirawski J, & Six S (1994) Oxygen regulated gene expression in facultatively anaerobic bacteria. *Antonie Van Leeuwenhoek* 66(1-3):3-22.
5. Schmitz RA DR, Deppenmeier U, Gottschalk G (2006) The anaerobic way of life. *The Prokaryotes*, (Springer), pp 86-101.
6. Hallmann C & Summons R (2014) Paleobiological clues to early atmospheric evolution. *Treatise on geochemistry, 2nd Ed*, (Elsevier), pp 139-155.
7. Canfield DE, Habicht KS, & Thamdrup B (2000) The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* 288(5466):658-661.
8. Madigan M, Martinko J, & Parker J (2000) *Brock biology of microorganisms* (Prentice-Hall) 9th ed Ed.
9. Sheridan PP, Freeman KH, & Brenchley JE (2003) Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiology Journal* 20(1):1-14.
10. Brochier-Armanet C, Talla E, & Gribaldo S (2009) The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* 26(2):285-297.
11. Boucher Y, *et al.* (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283-328.
12. Stolper DA, Revsbech NP, & Canfield DE (2010) Aerobic growth at nanomolar oxygen concentrations. *Proc Natl Acad Sci U S A* 107(44):18755-18760.
13. Muyzer G & Stams AJ (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nat Rev Microbiol* 6(6):441-454.
14. Lagunas R (1982) What is meant by facultative anaerobes? *Biochem Educ* 10(4):141-142.
15. Fuchs G (2008) Anaerobic metabolism of aromatic compounds. *Ann N Y Acad Sci* 1125:82-99.

16. Shinoda Y, *et al.* (2004) Aerobic and anaerobic toluene degradation by a newly isolated denitrifying bacterium, *Thauera* sp. strain DNT-1. *Appl Environ Microbiol* 70(3):1385-1392.
17. Heider J, *et al.* (1998) Differential induction of enzymes involved in anaerobic metabolism of aromatic compounds in the denitrifying bacterium *Thauera aromatica*. *Arch Microbiol* 170(2):120-131.
18. Verfurth K, Pierik AJ, Leutwein C, Zorn S, & Heider J (2004) Substrate specificities and electron paramagnetic resonance properties of benzylsuccinate synthases in anaerobic toluene and *m*-xylene metabolism. *Arch Microbiol* 181(2):155-162.
19. Heider J, Spormann AM, Beller HR, & Widdel F (1998) Anaerobic bacterial metabolism of hydrocarbons. *FEMS microbiology reviews* 22(5):459-473.
20. Begley TP (2006) Cofactor biosynthesis: an organic chemist's treasure trove. *Nat Prod Rep* 23(1):15-25.
21. Reichard P (1993) From RNA to DNA, why so many ribonucleotide reductases? *Science* 260(5115):1773-1777.
22. Stubbe J (2000) Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr Opin Struct Biol* 10(6):731-736.
23. Lundin D, Gribaldo S, Torrents E, Sjöberg BM, & Poole AM (2010) Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC Evol Biol* 10:383.
24. Torrents E, Aloy P, Gibert I, & Rodriguez-Trelles F (2002) Ribonucleotide reductases: divergent evolution of an ancient enzyme. *J Mol Evol* 55(2):138-152.
25. Jordan A, *et al.* (1999) Ribonucleotide reduction in *Pseudomonas* species: simultaneous presence of active enzymes from different classes. *J Bacteriol* 181(13):3974-3980.
26. Fontecave M, Eliasson R, & Reichard P (1989) Enzymatic regulation of the radical content of the small subunit of *Escherichia coli* ribonucleotide reductase involving reduction of its redox centers. *J Biol Chem* 264(16):9164-9170.
27. Borovok I, Gorovitz B, Schreiber R, Aharonowitz Y, & Cohen G (2006) Coenzyme B12 controls transcription of the *Streptomyces* class Ia ribonucleotide reductase *nrdABS* operon via a riboswitch mechanism. *J Bacteriol* 188(7):2512-2520.
28. Martin JE & Imlay JA (2011) The alternative aerobic ribonucleotide reductase of *Escherichia coli*, NrdEF, is a manganese-dependent enzyme that enables cell replication during periods of iron starvation. *Mol Microbiol* 80(2):319-334.
29. Hioe J, Savasci G, Brand H, & Zipse H (2011) The stability of C α peptide radicals: why glyxyl radical enzymes? *Chemistry* 17(13):3781-3789.
30. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA. How the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.
31. Parast CV, Wong KK, Kozarich JW, Peisach J, & Magliozzo RS (1995) Electron paramagnetic resonance evidence for a cysteine-based radical in pyruvate formate-lyase inactivated with mercaptopyruvate. *Biochemistry* 34(17):5712-5717.
32. Parast CV, *et al.* (1995) Hydrogen exchange of the glyxyl radical of pyruvate formate-lyase is catalyzed by cysteine 419. *Biochemistry* 34(8):2393-2399.
33. Rauk A, Yu D, & Armstrong D (1998) Oxidative damage to and by cysteine in proteins: An ab initio study of the radical structures, CH, SH, and CC bond dissociation energies, and

- transition structures for H abstraction by thiyl radicals. *Journal of the American Chemical Society* 120(34):8848-8855.
34. Armstrong D, Yu D, & Rauk A (1996) Oxidative damage to the glycyl α -carbon site in proteins: an ab initio study of the CH bond dissociation energy and the reduction potential of the C-centered radical. *Canadian Journal of Chemistry* 74(6):1192-1199.
 35. Wei Y, *et al.* (2014) The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proc Natl Acad Sci U S A* 111(36):E3756-3765.
 36. Knappe J & Sawers G (1990) A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol Rev* 6(4):383-398.
 37. Knappe J, Blaschkowski HP, Grobner P, & Schmitt T (1974) Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. *Eur J Biochem* 50(1):253-263.
 38. Trchounian K, Poladyan A, Vassilian A, & Trchounian A (2012) Multiple and reversible hydrogenases for hydrogen production by *Escherichia coli*: dependence on fermentation substrate, pH and the F(0)F(1)-ATPase. *Crit Rev Biochem Mol Biol* 47(3):236-249.
 39. Thauer RK, Kirchner FH, & Jungermann KA (1972) Properties and function of the pyruvate-formate-lyase reaction in clostridia. *Eur J Biochem* 27(2):282-290.
 40. Leuthner B, *et al.* (1998) Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* 28(3):615-628.
 41. Leuthner B & Heider J (2000) Anaerobic toluene catabolism of *Thauera aromatica*: the *bbs* operon codes for enzymes of beta oxidation of the intermediate benzylsuccinate. *J Bacteriol* 182(2):272-277.
 42. Annweiler E, *et al.* (2000) Anaerobic degradation of 2-methylnaphthalene by a sulfate-reducing enrichment culture. *Appl Environ Microbiol* 66(12):5329-5333.
 43. Muller JA, Galushko AS, Kappler A, & Schink B (2001) Initiation of anaerobic degradation of *p*-cresol by formation of 4-hydroxybenzylsuccinate in *Desulfobacterium cetonicum*. *J Bacteriol* 183(2):752-757.
 44. Muller JA, Galushko AS, Kappler A, & Schink B (1999) Anaerobic degradation of *m*-cresol by *Desulfobacterium cetonicum* is initiated by formation of 3-hydroxybenzylsuccinate. *Arch Microbiol* 172(5):287-294.
 45. Callaghan AV, Wawrik B, Ni Chadhain SM, Young LY, & Zylstra GJ (2008) Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* 366(1):142-148.
 46. Li L, *et al.* (2009) Subunit structure of benzylsuccinate synthase. *Biochemistry* 48(6):1284-1292.
 47. Zengler K, Heider J, Rossello-Mora R, & Widdel F (1999) Phototrophic utilization of toluene under anoxic conditions by a new strain of *Blastochloris sulfoviridis*. *Arch Microbiol* 172(4):204-212.
 48. Kane SR, Beller HR, Legler TC, & Anderson RT (2002) Biochemical and genetic evidence of benzylsuccinate synthase in toluene-degrading, ferric iron-reducing *Geobacter metallireducens*. *Biodegradation* 13(2):149-154.
 49. Beller HR & Edwards EA (2000) Anaerobic toluene activation by benzylsuccinate synthase in a highly enriched methanogenic culture. *Appl Environ Microbiol* 66(12):5503-5505.
 50. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.

51. Martinez-Del Campo A, *et al.* (2015) Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6(2).
52. Craciun S, Marks JA, & Balskus EP (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycy radical enzymes. *ACS Chem Biol* 9(7):1408-1413.
53. Raynaud C, Sarcabal P, Meynial-Salles I, Croux C, & Soucaille P (2003) Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci U S A* 100(9):5010-5015.
54. Jorda J, Lopez D, Wheatley NM, & Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179-195.
55. O'Brien JR, *et al.* (2004) Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.
56. D'Ari L & Barker HA (1985) *p*-Cresol formation by cell-free extracts of *Clostridium difficile*. *Arch Microbiol* 143(3):311-312.
57. Andrei PI, Pierik AJ, Zauner S, Andrei-Selmer LC, & Selmer T (2004) Subunit composition of the glycy radical enzyme *p*-hydroxyphenylacetate decarboxylase. A small subunit, HpdC, is essential for catalytic activity. *Eur J Biochem* 271(11):2225-2230.
58. Dawson LF, Stabler RA, & Wren BW (2008) Assessing the role of *p*-cresol tolerance in *Clostridium difficile*. *J Med Microbiol* 57(Pt 6):745-749.
59. Blaser M (2006) Activation and regulation of the 4-hydroxyphenylacetate decarboxylase system from *Clostridium difficile*. Ph. D. Thesis (Philipps-Universität Marburg, Marburg, Germany).
60. Martins BM, *et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycy radical enzyme. *J Am Chem Soc* 133(37):14666-14674.
61. Yu L (2006) Two novel glycy radical decarboxylase systems from *Clostridium scatologenes* and *Tannerella forsythensis*. Ph. D. Thesis (Philipps-Universität Marburg, Marburg, Germany).
62. Mulliez E, Ollagnier S, Fontecave M, Eliasson R, & Reichard P (1995) Formate is the hydrogen donor for the anaerobic ribonucleotide reductase from *Escherichia coli*. *Proc Natl Acad Sci U S A* 92(19):8759-8762.
63. Dowling DP, Croft AK, & Drennan CL (2012) Radical use of Rossmann and TIM barrel architectures for controlling coenzyme B12 chemistry. *Annu Rev Biophys* 41:403-427.
64. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.
65. Chatterjee A, *et al.* (2008) Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily. *Nat Chem Biol* 4(12):758-765.
66. Shibata N, Higuchi Y, & Toraya T (2011) How coenzyme B12-dependent ethanolamine ammonia-lyase deals with both enantiomers of 2-amino-1-propanol as substrates: structure-based rationalization. *Biochemistry* 50(4):591-598.
67. Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, & Drennan CL (2004) Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science* 303(5654):76-79.
68. Uhlin U & Eklund H (1994) Structure of ribonucleotide reductase protein R1. *Nature* 370(6490):533-539.

69. Becker A, *et al.* (1999) Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Biol* 6(10):969-975.
70. Larsson KM, Andersson J, Sjöberg BM, Nordlund P, & Logan DT (2001) Structural basis for allosteric substrate specificity regulation in anaerobic ribonucleotide reductases. *Structure* 9(8):739-750.
71. Hinckley GT & Frey PA (2006) Cofactor dependence of reduction potentials for [4Fe-4S]^{2+/1+} in lysine 2,3-aminomutase. *Biochemistry* 45(10):3219-3225.
72. Wang SC & Frey PA (2007) Binding energy in the one-electron reductive cleavage of S-adenosylmethionine in lysine 2,3-aminomutase, a radical SAM enzyme. *Biochemistry* 46(45):12889-12895.
73. Wetmore SD, Smith DM, Golding BT, & Radom L (2001) Interconversion of (S)-glutamate and (2S,3S)-3-methylaspartate: a distinctive B(12)-dependent carbon-skeleton rearrangement. *J Am Chem Soc* 123(33):7963-7972.
74. Walsby CJ, *et al.* (2002) Electron-nuclear double resonance spectroscopic evidence that S-adenosylmethionine binds in contact with the catalytically active [4Fe-4S](+) cluster of pyruvate formate-lyase activating enzyme. *J Am Chem Soc* 124(12):3143-3151.
75. Dey A, *et al.* (2011) S K-edge XAS and DFT calculations on SAM dependent pyruvate formate-lyase activating enzyme: nature of interaction between the Fe₄S₄ cluster and SAM and its role in reactivity. *J Am Chem Soc* 133(46):18656-18662.
76. Duschene KS, Veneziano SE, Silver SC, & Broderick JB (2009) Control of radical chemistry in the AdoMet radical enzymes. *Curr Opin Chem Biol* 13(1):74-83.
77. Vey JL, *et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.
78. Duee ED, *et al.* (1994) Refined crystal structure of the 2[4Fe-4S] ferredoxin from *Clostridium acidurici* at 1.84-Å resolution. *J Mol Biol* 243(4):683-695.
79. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.
80. Dowling DP, *et al.* (2014) Radical SAM enzyme QueE defines a new minimal core fold and metal-dependent mechanism. *Nat Chem Biol* 10(2):106-112.
81. Frey M, Rothe M, Wagner AF, & Knappe J (1994) Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. *J Biol Chem* 269(17):12432-12437.
82. Nnyepi MR, Peng Y, & Broderick JB (2007) Inactivation of *E. coli* pyruvate formate-lyase: role of AdhE and small molecules. *Arch Biochem Biophys* 459(1):1-9.
83. Peng Y, Veneziano SE, Gillispie GD, & Broderick JB (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J Biol Chem* 285(35):27224-27231.
84. Crain AV & Broderick JB (2014) Pyruvate formate-lyase and its activation by pyruvate formate-lyase activating enzyme. *J Biol Chem* 289(9):5723-5729.
85. Mulliez E, Padovani D, Atta M, Alcouffe C, & Fontecave M (2001) Activation of class III ribonucleotide reductase by flavodoxin: a protein radical-driven electron transfer to the iron-sulfur center. *Biochemistry* 40(12):3730-3736.
86. Unkrig V, Neugebauer FA, & Knappe J (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analogue hypophosphite. *Eur J Biochem* 184(3):723-728.

87. Selvaraj B, Pierik AJ, Bill E, & Martins BM (2013) 4-Hydroxyphenylacetate decarboxylase activating enzyme catalyses a classical S-adenosylmethionine reductive cleavage reaction. *J Biol Inorg Chem* 18(6):633-643.
88. Funk MA, Judd ET, Marsh EN, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci U S A* 111(28):10161-10166.
89. Yu L, Blaser M, Andrei PI, Pierik AJ, & Selmer T (2006) 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* 45(31):9584-9592.
90. Kessler D, Leibrecht I, & Knappe J (1991) Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by adhE. *FEBS Lett* 281(1-2):59-63.
91. Wei Y, *et al.* (2014) A chemically competent thiosulfuranyl radical on the *Escherichia coli* class III ribonucleotide reductase. *J Am Chem Soc* 136(25):9001-9013.

Chapter II

Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity

This chapter has been published in a similar form: Funk, M. A., Judd, E. T., Marsh, E. N. G., Elliott, S. J., & Drennan, C. L. (2014). Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proceedings of the National Academy of Sciences*, 111(28), 10161-10166.

This chapter was written with Prof. Catherine L. Drennan. The electrochemical experiments in this chapter were performed by Dr. Evan T. Judd in collaboration with Prof. Sean J. Elliott (Boston University). Prof. E. Neil G. Marsh (University of Michigan) provided constructs and reagents.

II.A Summary

Anaerobic degradation of the environmental pollutant toluene is initiated by the glycyl radical enzyme benzylsuccinate synthase (BSS), which catalyzes the radical addition of toluene to fumarate, forming benzylsuccinate. We have determined crystal structures of the catalytic α subunit of BSS with its accessory subunits β and γ , which both bind a [4Fe-4S] cluster and are essential for BSS activity in vivo. We find that BSS α has the common glycyl radical enzyme fold, a 10-stranded β/α barrel that surrounds the glycyl radical cofactor and active site. Both accessory subunits β and γ display folds related to high potential iron-sulfur proteins but differ substantially from each other in how they interact with the α subunit. BSS γ binds distal to the active site, burying a hydrophobic region of BSS α , whereas BSS β binds to a hydrophilic surface of BSS α that is proximal to the active site. To further investigate the function of BSS β , we determined the structure of a BSS $\alpha\gamma$ complex. Remarkably, we find that the barrel partially opens, allowing the C-terminal region of BSS α that houses the glycyl radical to shift within the barrel toward an exit pathway. The structural changes that we observe in the BSS $\alpha\gamma$ complex center around the crucial glycyl radical domain, thus suggesting a role for BSS β in modulating the conformational dynamics required for enzyme activity. Accompanying proteolysis experiments support these structural observations.

II.B Introduction

Microbial degradation of hydrocarbons is a crucial component of the restoration of ecosystems after an influx of natural or man-made pollution. Aromatic hydrocarbons, primarily toluene and its derivatives, are some of the most abundant molecules present in groundwater pollution and thus are a prominent target for bioremediation efforts. The persistence of these molecules is due to their solubility and inertness: compounds such as toluene, phenol, and ethylbenzene contain no easily oxidized carbons and dearomatization of these compounds requires strong oxidants or reductants. Nevertheless, microorganisms have developed diverse strategies to activate and degrade these compounds under both aerobic and anaerobic conditions (1).

Benzylsuccinate synthase (BSS) catalyzes the first step in anaerobic toluene degradation, the generation of (*R*)-benzylsuccinate from toluene and fumarate (Figure II.1A) (2). This unusual reaction is thought to proceed through radical mediated C-C bond formation (3-6), with the

radical source being an enzyme-bound glycyl (Gly) radical cofactor (7) that transiently forms a catalytically essential enzyme-bound thiyl (Cys) radical. Close family members of BSS, henceforth called the X-succinate synthase group, catalyze conjugation of fumarate and 2-methylnaphthalene, *p*-cresol, and hydrocarbons including *n*-hexane (8, 9). The X-succinate synthase group is a member of the larger Gly radical enzyme (GRE) family (Figure II.2), which includes class III ribonucleotide reductase (anRNR), pyruvate formate lyase (PFL), glycerol dehydratase (GDH), and 4-hydroxyphenylacetate decarboxylase (HPD) (10). These enzymes share structural features including a 10-stranded β/α barrel, a Cys loop containing the cysteine residue essential for catalytic hydrogen atom transfer, and a Gly radical loop housed within the C-terminal Gly radical domain that functions in radical storage.

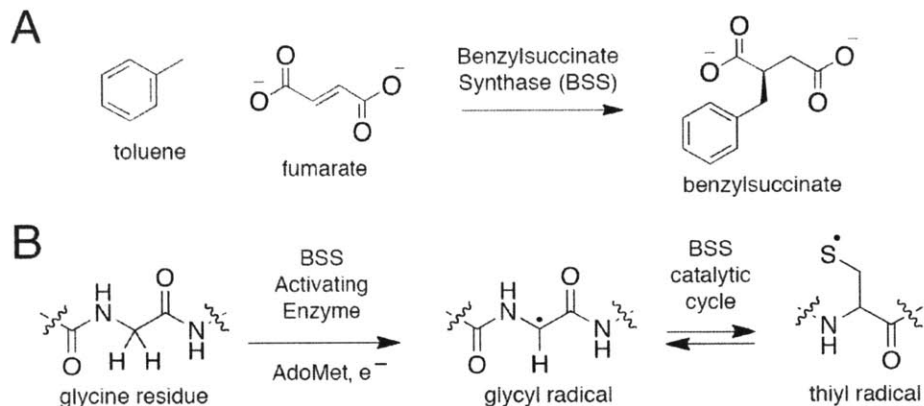


Figure II.1. Benzylsuccinate synthase (BSS) activation and reaction. A) Active BSS catalyzes C-C bond formation between toluene and fumarate. **B)** BSS-AE generates a Gly radical on BSS.

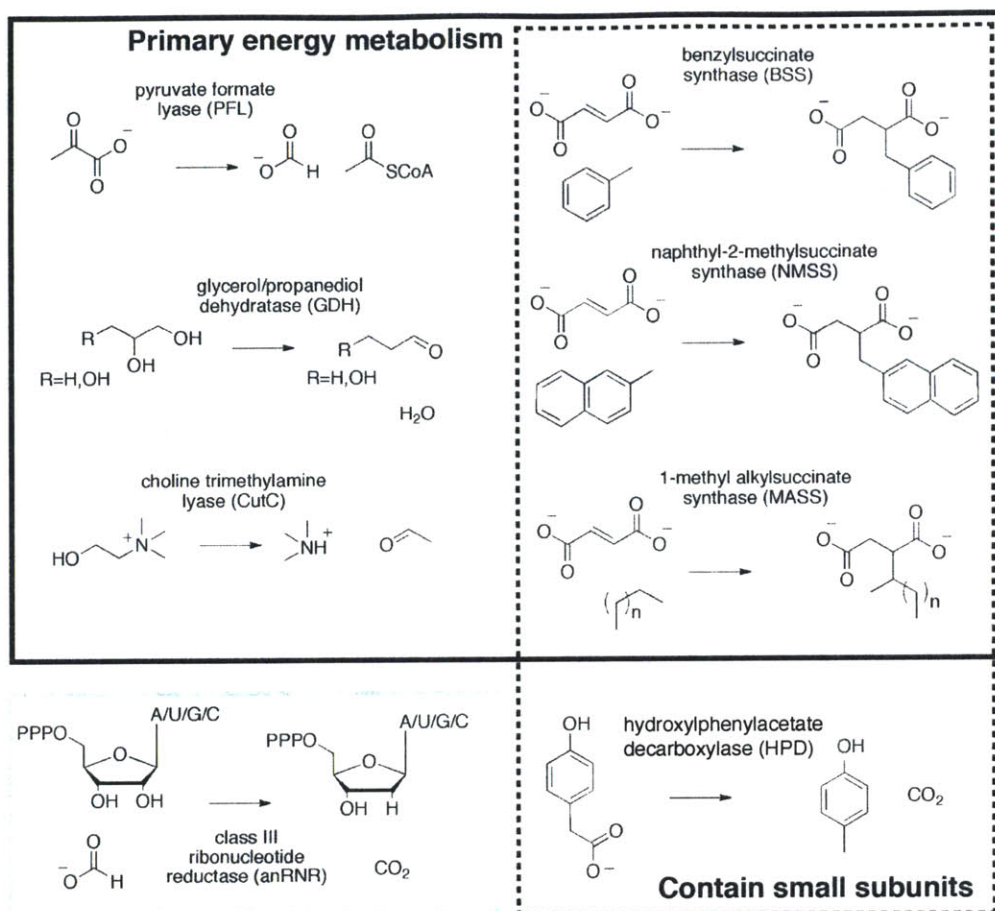


Figure II.2. Glycyl radical enzymes (GREs) enable challenging chemistry to be performed anaerobically. Reactions catalyzed by GREs (10), including the recently characterized choline trimethylamine lyase, CutC (11).

The Gly radical is generated post-translationally through direct hydrogen atom abstraction by a GRE activating enzyme (AE) (12), a member of the radical *S*-adenosyl-L-methionine superfamily (Figure II.3) (13). A structure of an AE bound to a peptide mimic of the GRE Gly radical domain (14) shows that the Gly radical domain must be extended away from the GRE to fit into the AE active site. However, once activated, the Gly radical domain must return to the inside of the barrel where the Gly radical can undergo hydrogen atom transfer with the catalytic Cys during each round of turnover, and where it is sequestered away from oxidants and reductants that could quench the radical. Consistent with this idea, all structural data thus far on GREs show the Gly radical domain buried deep within the GRE core adjacent to the Cys loop. Thus, two states of any GRE must exist, one with the Gly radical domain extended out of the protein core for post-translational modification by the AE, and one closed for catalysis. Using CD

spectroscopy and analysis of enzyme activity, the Gly radical domain of PFL was shown to undergo structural changes when binding to its activase PFL-AE (15), with PFL-AE either catalyzing egress of the Gly radical domain, or shifting the equilibrium of closed and open states. Despite these glimpses of the activation process, the mechanism by which this large conformational change is initiated, reversed, and regulated remains enigmatic.

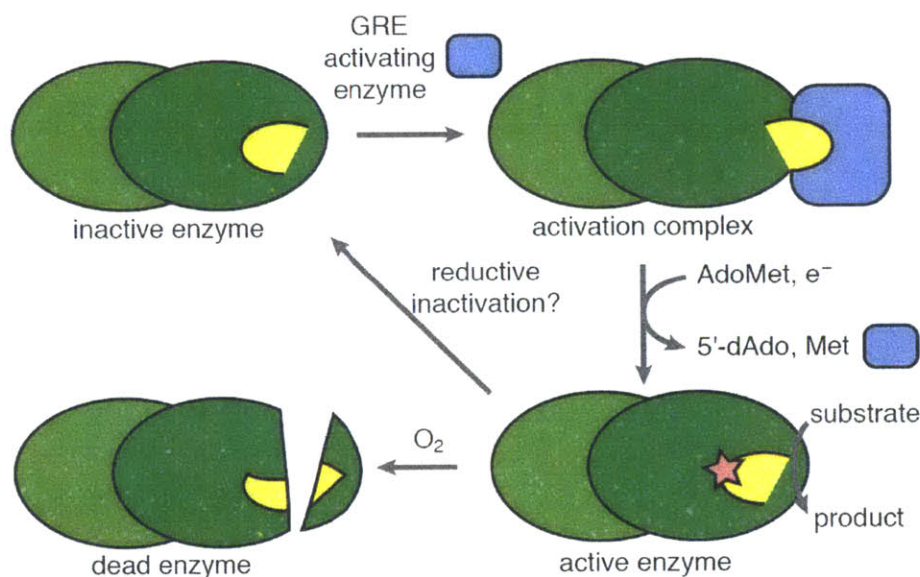


Figure II.3. GREs are post-translationally activated. GREs are produced as inactive proenzymes containing an absolutely conserved glycine residue on a small C-terminal Gly radical domain (yellow). The inactive enzyme is primarily folded in a closed conformation—the resting state—but is thought to be in equilibrium with an open state in which the Gly radical domain is exposed. In this open state, an activation complex can be formed with the GRE activating enzyme (GRE-AE, blue). The GRE-AE generates the Gly radical through reductive cleavage of *S*-adenosyl-L-methionine (AdoMet), transiently forming a 5'-deoxyadenosyl radical that abstracts a hydrogen atom from the Gly C α , and ultimately forming the Gly radical, methionine (Met) and 5'-deoxyadenosine (5'-dAdo). The activated enzyme returns to a closed conformation and can catalyze turnover. Irreversible inactivation of the enzyme occurs through oxidative quenching of the Gly radical and subsequent cleavage of the protein backbone. Various reductants can reversibly reduce the Gly radical back to glycine, although the *in vivo* relevance of this inactivation is not well established.

BSS is distinguished among characterized GREs as one of the few that is composed of multiple subunits; a stable ($\alpha\beta\gamma$)₂ heterohexamer can be isolated from cultured *Thauera aromatica* strain T1 grown on toluene (16) or expressed recombinantly in *Escherichia coli* (17). The 98-kDa α subunit contains the catalytic machinery and forms a homodimeric complex, a common feature of GREs. BSS β and γ are much smaller, 9 and 7 kDa, respectively, and have

been investigated on their own (17) and in complex with the α subunit (17, 18). Electron paramagnetic resonance, Mössbauer, and UV-visible spectroscopy from these studies indicate that BSS β and $-\gamma$ each contain a [4Fe-4S] cluster ligated by four conserved cysteines in these proteins. In both subunits the cluster is a [4Fe-4S]²⁺ species that can be reduced to form the [4Fe-4S]¹⁺ species after treatment with dithionite. The only structural information available for multi-subunit GREs comes from HPD. The structure of the catalytic subunit of HPD, HpdB, in complex with its small subunit, HpdC, was recently solved, revealing that HpdC is bound at the protein surface away from the active site and the Gly radical domain (19).

In this study, we investigate the structure of BSS, the mechanism of post-translational activation, and the role of BSS β and $-\gamma$ in this system. Although *in vivo* experiments have demonstrated both subunits are essential for *in vivo* toluene utilization (20), their functions in the system have not been established. We have determined structures of two complexes, BSS $\alpha\beta\gamma$ and BSS $\alpha\gamma$, which reveal structural and dynamic changes in the catalytic α subunit in response to binding of BSS β . These results suggest BSS β acts as a regulator of activation and may additionally be involved in regulating access to the enzyme's active site. Excitingly, we have observed two different states of the Gly radical domain in BSS. The movements seen in these structures are consistent with large conformational changes occurring during the activation of GREs by their cognate AE.

II.C Results

Overall structure of BSS

We crystallized and determined the structure of BSS from *T. aromatica* strain T1, a denitrifying bacterium. BSS α , $-\beta$, and $-\gamma$ were co-expressed in *E. coli* and purified as an ($\alpha\beta\gamma$)₂ complex. The catalytic α subunit adopts the core fold observed in other GREs (Figure II.4 and II.5): a 10-stranded $\beta\alpha$ barrel with five strands oriented in one direction and the other five in the opposite direction. The barrel encloses two loops that harbor the catalytic Gly and Cys residues. The Cys loop contains the catalytic Cys493 residue and connects two β strands that insert through the center of the $\beta\alpha$ barrel. The four backbone amide nitrogens from residues Val491 to Ser495 point inward to form a shallow pocket that coordinates a chloride anion (Figure II.6A). This site is occupied by the carboxylate of a neighboring Glu in HPD and GDH, but Ser or Ala replace this

Glu in X-succinate synthase enzymes. Thus it is tempting to speculate that the common dicarboxylate substrate of X-succinate synthases, fumarate, could displace Cl^- and bind directly to the Cys loop, mimicking the Glu interaction found in HPD and GDH. At the opposite end of the pocket (Figure II.6B), Arg508 is absolutely conserved among X-succinate synthases, and could play a role in binding fumarate as well. One wall of the active site cavity is formed entirely of hydrophobic residues, providing a surface for toluene binding. Adjacent to the Cys loop is the Gly loop, found at the innermost tip of the C-terminal Gly radical domain, the key structural element believed to undergo a large conformational change during activation.

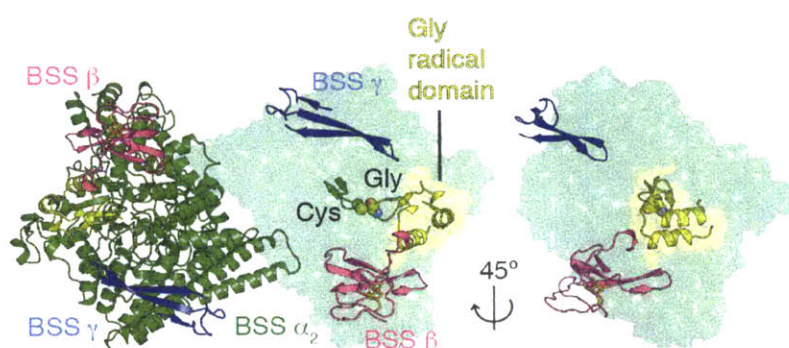


Figure II.4. BSS(αβγ)₂ is a heterohexamer. The BSSα₂ dimer (green) is shown with the left monomer in ribbons and the right in surface representation (also see Figure II.S2). The Gly and Cys loops meet in the center of the protein; the catalytic Gly and Cys are shown in spheres. BSSβ (magenta) contains one [4Fe-4S] cluster; BSSγ (blue) is lacking a [4Fe-4S] cluster in this structure. The Gly radical domain of BSSα (yellow) interacts with the C terminus of BSSβ, as shown in the side view.

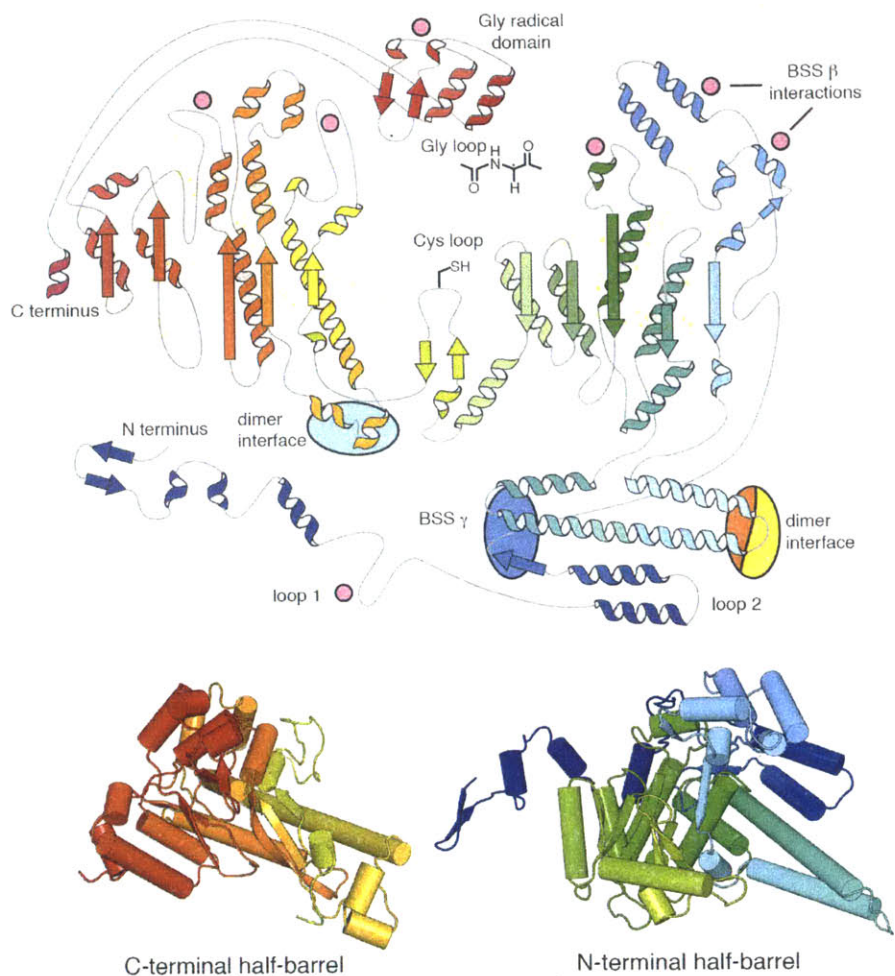


Figure II.5. Topology and ribbon diagrams for BSS. *Top*, secondary structure of BSS α is shown in the form of a topology diagram in rainbow colors from N (blue) to C (red) terminus. The core of the α subunit is composed of two $(\beta\alpha)_5$ half-barrels (shown at bottom), which sandwich the catalytic Gly and Cys loops in the interior of the protein. Surrounding the central β barrel are exterior α helices and a C-terminal helix bundle, both of which contribute to shielding of the active site and stabilization of the core. The dimer interface is primarily composed of two sites that contact each other: the end of a helix bundle near the N terminus (yellow/orange oval) and outside the C-terminal half-barrel (light blue oval). The C-terminal tail in BSS α is extended relative to other GREs by 10-15 residues. These residues form a short helix that rests in a groove on the outer surface of the enzyme, burying over 1200 \AA^2 of surface area. The N terminus is disordered in most structures of GREs, but has been observed in structures of pyruvate formate lyase wrapped around the protein through a series of extended loops and helices. The N terminus of BSS likewise lies in an extended conformation along the outside of the protein, primarily against the Gly radical domain and C terminus. BSS β contacts numerous regions of the protein at the top of the half-barrels, shown as pink circles. BSS γ contacts only a small region at the end of the dimer-forming helix bundle, shown as a dark blue oval. Loop 1 and 2 are labeled. *Bottom*, cartoon diagram of two $(\beta\alpha)_5$ half-barrels that make up the 10-stranded barrel of BSS.

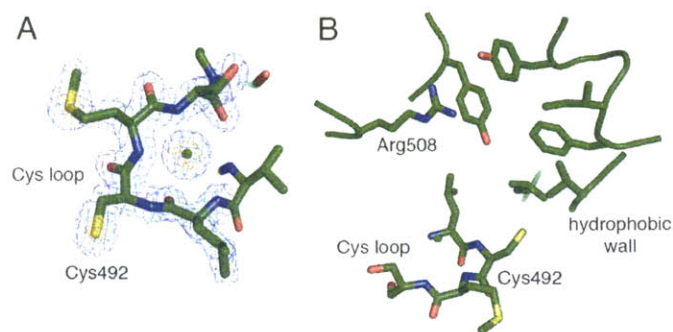


Figure II.6. Active site environment in BSS. A) Cys loop with a chloride binding site in the BSS $\alpha\beta\gamma$ structure. $2mF_o-DF_c$ density contoured at 1σ (blue mesh) and 5σ (orange mesh) is consistent with a chloride ion (green sphere) which coordinates the four amide nitrogens of residues 492-495 with Cl-N distances between 3.12-3.50 Å. B) The active site cavity contains the Cys loop, conserved Arg508, and a hydrophobic wall.

The structure of the BSS $\alpha\beta\gamma$ complex reveals two exterior binding sites for the small β and γ subunits (Figure II.4), with no changes to the core barrel architecture of BSS α , as compared to other GREs that do not have additional subunits. Within one $\alpha\beta\gamma$ unit, the average distance between BSS β and $-\gamma$ is 45 Å, with the closest contact distance being 13 Å. The subunits bind to different regions of BSS α in both primary and tertiary structure and have radically different binding interfaces. A number of unique structural features are present at the surface of BSS α that enable binding to the small subunits as discussed below.

Despite having no recognizable sequence homology to any protein of known structure, BSS β adopts a fold similar to that of a high potential iron-sulfur protein (HiPIP) (Figure II.7A). Three short β strands make a stable platform for the coordination of the [4Fe-4S] cluster by four Cys ligands. A 3_{10} helix abuts the cluster and contains two of the Cys thiolate ligands, and long loops at the N and C termini wrap around the cluster, shielding it from solvent on two faces. Comparison to a characterized HiPIP reveals that several hydrophobic residues that normally exclude solvent from the cluster environment and enable stabilization of the [4Fe-4S] $^{3+}$ state are missing in BSS β or replaced by hydrophilic residues (Figure II.8A). The region around Arg69 in particular is directly open to solvent molecules, and several waters are observed adjacent to the cluster, 3.7 Å from S_γ of Cys44. BSS β contains an extension of the basal β sheet by several residue pairs compared to HiPIPs, which provides a larger surface area for making contacts to the α subunit (Figure II.8B). Additionally, BSS β contains an extended hairpin loop and a C-terminal

3_{10} helix (Figure II.8A), both of which are involved in inter-subunit interactions as described below.

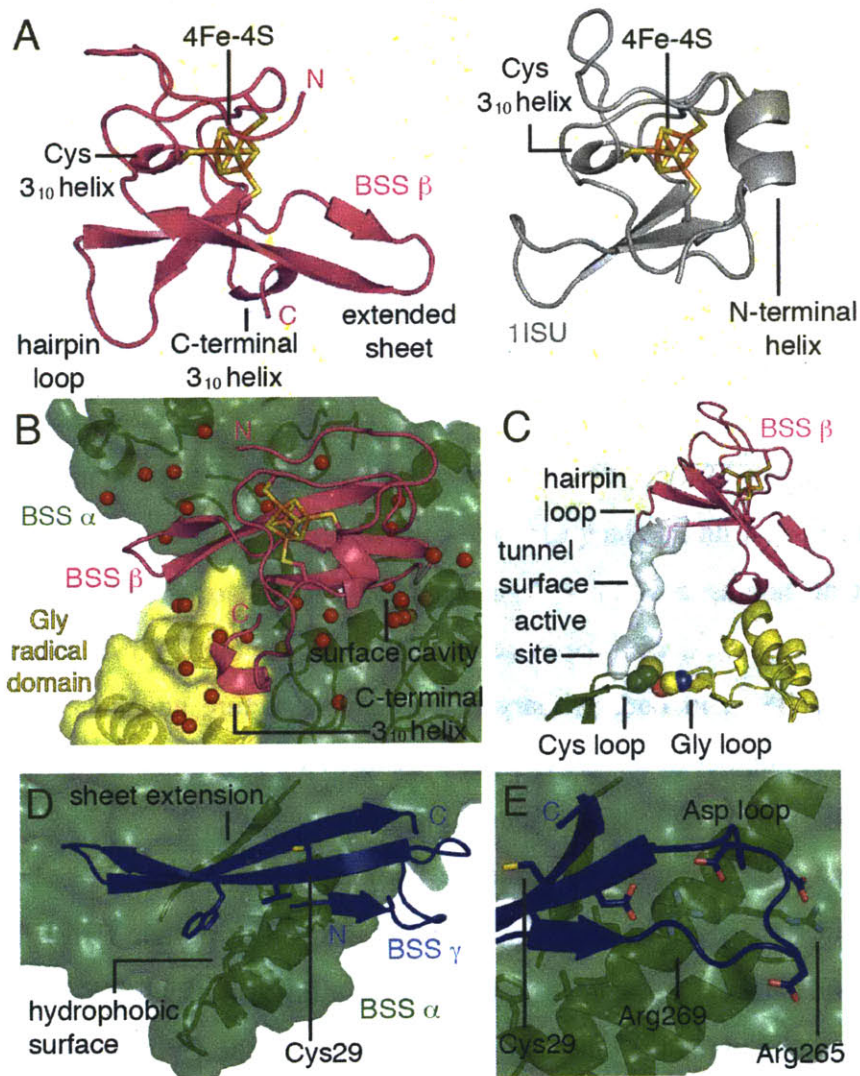


Figure II.7. BSS small subunits and subunit interfaces. **A)** Comparison of BSS β (magenta) to a model HiPIP (gray, PDB 1ISU), with fully Cys-liganded [4Fe-4S] clusters shown in sticks. The C terminus and last visible residue at the N terminus are labeled for BSS β . **B)** BSS β binds a broad surface on BSS α (green) that includes regions of the Gly radical domain (yellow). A small cavity on the surface of BSS α is occupied by a β hairpin loop from BSS β . Waters connecting the two subunits by hydrogen bonds are shown as red spheres. **C)** A putative substrate access tunnel is shown in a white surface extending up from the catalytic Cys loop toward the enzyme surface. **D)** BSS γ (blue) binds BSS α at a hydrophobic surface near the BSS α dimer interface. **E)** Additional ionic interactions are present between BSS γ and BSS α .

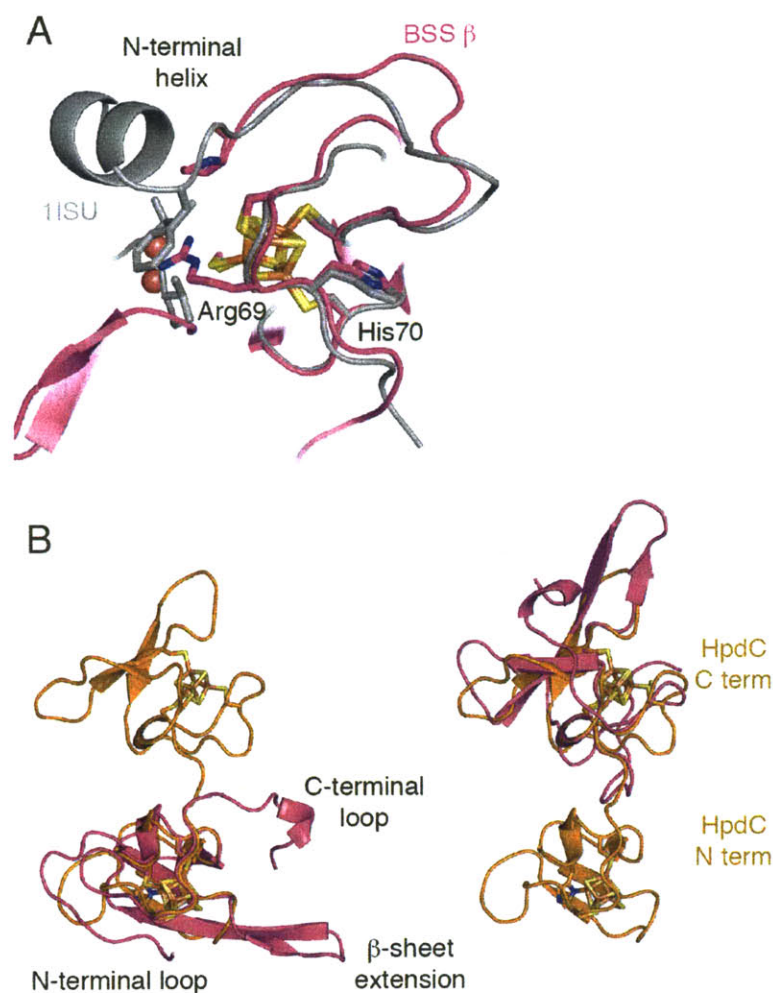


Figure II.8. Comparisons of high potential iron-sulfur protein (HiPIP) folds. A) Comparison of BSS β (magenta) to a representative HiPIP (1ISU, gray) showing some of the features that lead to a more hydrophilic environment around the [4Fe-4S] cluster in BSS. In particular, Arg69 and His70 of BSS β (shown in sticks) replace conserved aromatic and hydrophobic residues in the HiPIP, creating a water-filled pocket (red spheres). **B)** BSS β (magenta) overlaid on the N-terminal domain (left) and with the C-terminal domain (right) of the HPD small subunit HpdC (orange). The superposition of the two subunits were calculated using C α positions of structurally conserved residues (typically the cluster-ligating residues and their immediate neighbors), yielding RMSDs of 0.78 and 0.77 for the N- and C-terminal domains, respectively.

BSS β also resembles the single small subunit of HPD, which is known as the HpdC or HPD γ in that system (Figure II.8B). HpdC contains two small, concatenated HiPIP domains connected by a short linker. BSS β superposes well with both of these domains (C α RMSD of 0.78 and 0.77 for the N- and C-terminal domains, respectively, based on a core excluding unshared regions), although neither contain the β sheet extension or the long N-terminal loop,

which shield the cluster in BSS β . The [4Fe-4S] cluster in BSS β is therefore intermediate in terms of sequestration from solvent between the less buried clusters present in HpdC, and the more buried clusters of HiPIPs. The C-terminal region is also unique to BSS β but does not contribute to cluster shielding.

Although BSS γ is clearly present in the structure of the BSS $\alpha\beta\gamma$ complex, the N and C termini are disordered, and there is no bound [4Fe-4S] cluster observable in the electron density. A close symmetry mate precludes packing of the intact subunit in both of the two crystal forms we have obtained (Figure II.9A-B). BSS γ is only 19% identical to BSS β , but they share the basic HiPIP fold along with the extended β sheet and hairpin loop (Figure II.9C). Modeling of the disordered residues in BSS γ based on the structure of BSS β suggests the cluster in BSS γ would be more exposed to solvent than in BSS β (Figure II.9D).

To investigate whether BSS β and $-\gamma$ display typical redox properties of HiPIP proteins, we performed protein film voltammetry (PFV) on the BSS $\alpha\beta\gamma$ and $-\alpha\gamma$ complexes (Figure II.10A-B). Consistent with the less hydrophobic environment observed for the [4Fe-4S] cluster of BSS β , a redox potential of ~ 80 mV is observed, which is in the lower end of the range for the HiPIP +3/+2 couple (50-500 mV, average ~ 300 mV) (21). BSS β and $-\gamma$ have similar potentials, as evidenced by the small differences between the voltammograms of BSS $\alpha\beta\gamma$ and $-\alpha\gamma$ complexes, such that the contributions of the two [4Fe-4S] clusters cannot be readily distinguished in the case of BSS $\alpha\beta\gamma$. Unlike true HiPIP proteins (21), the [4Fe-4S] clusters of both BSS $\alpha\beta\gamma$ and $-\alpha\gamma$ engage in redox chemistry that is coupled to a single proton over a pH range of 6 to 9 (Figure II.10C). However, the [4Fe-4S] clusters in the BSS small subunits are resistant to degradation by oxygen (Figure II.11), a characteristic shared with HiPIP proteins (22).

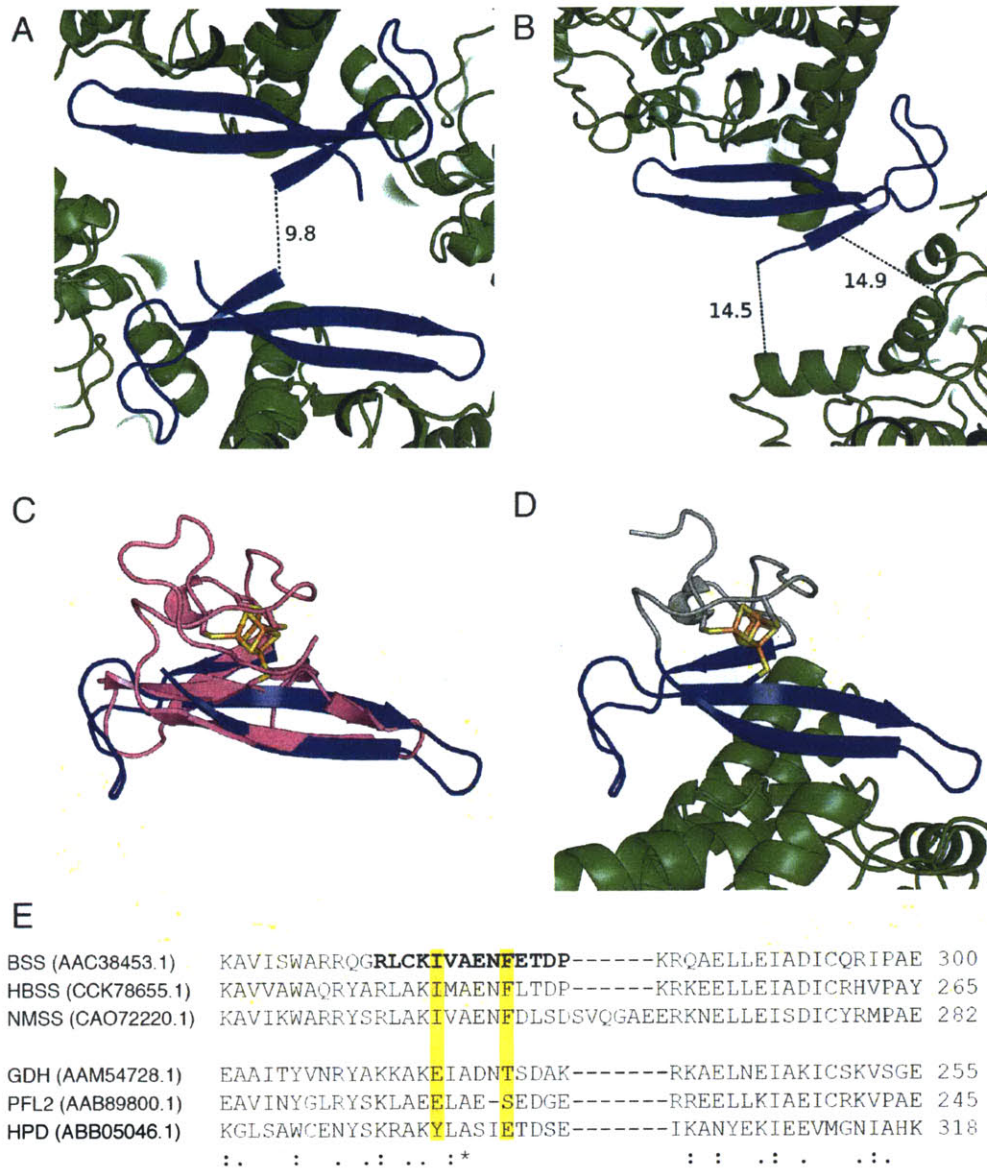


Figure II.9. The BSS γ [4Fe-4S] cluster site is disordered in both crystal forms. A) Crystal packing interactions around BSS γ in the BSS $\alpha\beta\gamma$; **B)** crystal packing around BSS γ in the $\alpha\gamma$ complexes; distances shown are in Å. **C)** Alignment of BSS β (magenta) and BSS γ (blue) based on Ca atoms of residues within the core β sheet. The Ca RMSD for the aligned residues is 1.16 Å. **D)** Hypothetical model for the complete structure of BSS γ . Residues in gray were modeled manually using the structure of BSS β as a guide. BSS γ has 15 fewer residues at its N terminus than BSS β . In BSS β , some of these residues help shield its [4Fe-4S] cluster, and as a result, the cluster in BSS γ is likely more solvent exposed. **E)** Sequence alignment of residues of BSS α to residues of other X-succinate synthase family members, highlighting the region at the BSS α - γ interface. Residues in bold denote the BSS α residues that contact BSS γ . Ile273 and Phe278 (yellow) are crucial to the hydrophobic contacts which dominate the α - γ interface and are conserved only in X-succinate synthase family members. GenBank accession numbers are

given for the selected sequences. BSS, 4-hydroxy-benzylsuccinate synthase (HBSS) and naphthyl-2-methyl-succinate synthase (NMSS) have a γ subunit present in the BSS operon. Glycerol dehydratase (GDH) and pyruvate formate lyase-like protein 2 (PFL2) contain a similar structural motif to that seen in BSS but do not bind a γ subunit. The HpdC subunit in the HPD system binds HpdB at a different site with different contacts from those seen in BSS.

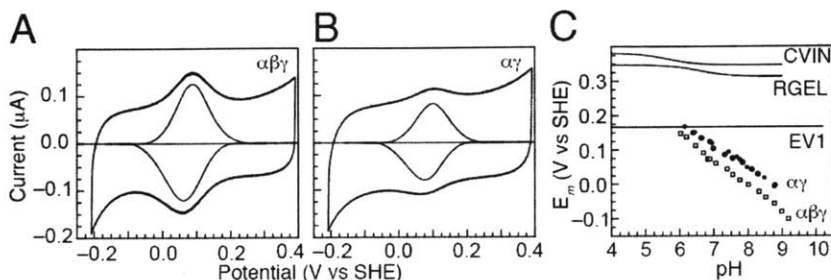


Figure II.10. BSS protein film voltammetry. **A)** Voltammetric response of BSS $\alpha\beta\gamma$. $E_m = 76$ mV. **B)** Voltammetric response of BSS $\alpha\gamma$. $E_m = 89$ mV. Raw cyclic voltammograms (outer) and curves with the background electrode capacitance corrected (inner) are shown. **C)** Dependence of E_m on pH for BSS $\alpha\beta\gamma$ and BSS $\alpha\gamma$ compared to the pH dependence of the E_m of *Chromatium vinosum* HiPIP (CVIN), the *Rhodocyclus gelatinosus* HiPIP (RGEL) and the *Ectothiorhodospira vacuolata iso-2* HiPIP (EV1) as determined previously (21).

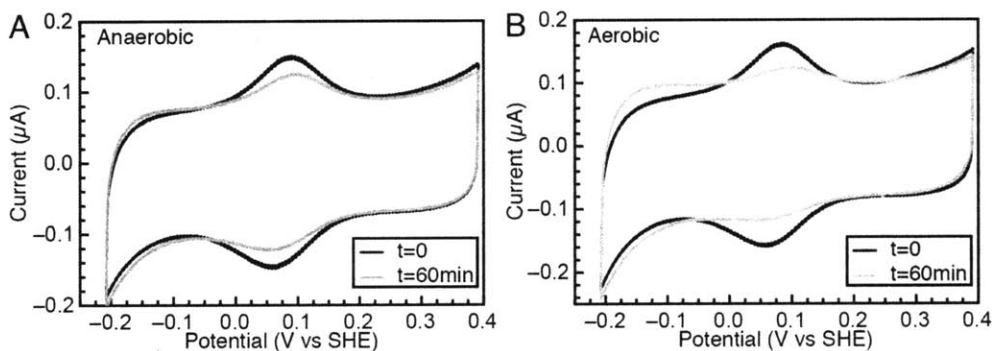


Figure II.11. Stability of BSS films under aerobic and anaerobic conditions. **A)** Representative BSS signal after initial film generation on 3-MPA/Au electrodes and after a 60 min incubation in deoxygenated buffer in an anaerobic environment. Oxidative signal area was $54 \pm 11\%$ of the initial ($t=0$) value after one hour. **B)** Representative BSS signal after initial film generation on 3-MPA/Au electrodes and after a 60 min incubation in oxygenated buffer in an aerobic environment. First and last scans were recorded in anaerobic environments and deoxygenated buffer to prevent oxygen interference with PFV experiments. Oxidative signal area was $36 \pm 4\%$ its initial ($t=0$) value after one hour. Scans recorded at pH 7, 60 mV/s, 20 °C. Signal loss under both anaerobic and aerobic conditions is presumed to be a mixed contribution of protein desorption and any cluster degradation. Although some loss is observed, a large amount of the original signal remains in the aerobically incubated sample despite two buffer exchanges during the course of the experiment, indicating good cluster stability.

BSS inter-subunit interactions

Previous studies on BSS have shown that expression and folding of BSS α depends on the co-expression of BSS γ but not BSS β (17). The structure of the complex reveals a mostly hydrophobic binding site for BSS γ (Figure II.7D), exposure of which might induce the aggregation or misfolding seen during expression of BSS α alone. The hydrophobic patch on BSS α is positioned at the end of the dimer-forming helices and includes Ile273 and Phe278 in BSS α . These residues are conserved within the X-succinate synthase group—all of which contain putative γ subunits—but not in more distant GRE families, including HPD (Figure II.9E). An inter-subunit β sheet is present with a single strand from BSS α and the two BSS γ strands that make the [4Fe-4S] cluster platform. The nature and arrangement of interactions leaves very little area exposed to solvent at the main interface. There are additional ionic interactions between a highly acidic loop in BSS γ and the main dimerization helix of BSS α , which contains two Arg residues (Figure II.7E).

In contrast to BSS α - γ interactions, the binding surface for BSS β on BSS α is not hydrophobic (Figure II.7B), nor is it confined to any single region in the primary sequence (Figure II.5). The primary interaction surface of BSS β , the side and underside of the β sheet, packs into a central indentation on BSS α , creating a hydrophilic interface with extensive polar side chain contacts and more than 20 buried water molecules. Several of the structural features that differentiate BSS β from traditional HiPIPs are observed to be important for binding to BSS α . In particular, the extended nature of the β sheet with a kink in the central β strand creates an unusually strong β sheet curvature that allows for numerous binding interactions with BSS α , including contacts to the Gly radical domain. Also, the C-terminal 3_{10} helix of BSS β makes multiple contacts with the Gly radical domain (Figure II.7B). Together, the C terminus and sheet of BSS β cover $\sim 220 \text{ \AA}^2$ (out of $\sim 1400 \text{ \AA}^2$) of solvent accessible surface at the Gly radical domain. With this binding mode, BSS β appears to stabilize the Gly radical domain inside the protein as any movement of the domain would require a change in conformation or dissociation of BSS β .

Part of the BSS β interaction surface includes a cavity proposed in GREs to be the physiological access route for substrate into the completely buried Cys loop at the active site (23). Using CAVER (24), a tunnel that is similar to ones described previously in GREs can be identified in BSS (Figure II.7C). However, in BSS, this putative substrate entryway is obstructed

by a hairpin loop from BSS β that inserts into the top of the cavity. Notably, this hairpin loop is another region that differs between BSS β and traditional HiPIP proteins. When bound to BSS α , the hairpin loop appears to be very stable and is perfectly structured to fit into the BSS α cavity and possibly gate substrate entry.

Global structural changes in the BSS $\alpha\gamma$ complex

In order to further investigate the role of BSS β in this system, we sought to obtain a structure lacking the subunit. We successfully expressed and purified the BSS $\alpha\gamma$ complex in *E. coli*, obtained crystals in a new crystal form, and determined the structure by molecular replacement using the structure of BSS α as a basis (Table 1). The overall structure of BSS α in complex with BSS γ alone retains many of the features of the complex with BSS β , especially in the N-terminal half-barrel and dimer interface, both of which are unchanged from BSS $\alpha\beta\gamma$ (Figure II.12A). However, the C-terminal half-barrel exhibits large crystal-to-crystal variation in the amount and quality of electron density around both core helices and peripheral loops and is consistently shifted relative to the BSS $\alpha\beta\gamma$ complex. Difference distance matrix analysis (Figure II.13) of the BSS $\alpha\beta\gamma$ and $-\alpha\gamma$ structures revealed a number of structural changes with distances unbiased by the choice of alignment strategy (25). With BSS β absent, BSS α expands in a clamshell-like motion by 5-7 Å, centering at helix 9 (residues 748-758), which is partially disordered, and strand 9 (residues 763-771) (Figure II.13). These motions open a gap in the center of the protein that allows the Gly radical domain to slip out of the active site by ~2 Å (Figure II.12B). Residues 55-65 of BSS α form a loop, henceforth referred to as loop 1, which contacts both the Gly radical domain and the C-terminal tail of BSS β (Figure II.12B). Loop 1 moves alongside the Gly radical domain away from the active site by ~3 Å in the BSS $\alpha\gamma$ structure relative to BSS $\alpha\beta\gamma$, possibly facilitating release of the Gly radical domain from the interior of the protein. At the active site, the C α -S γ distance between Gly829 and Cys493 lengthens in the $\alpha\gamma$ complex by 0.60 Å to 4.45 Å. Although the Gly radical domain remains within the barrel opening, the absence of BSS β removes a major obstacle from its exit pathway.

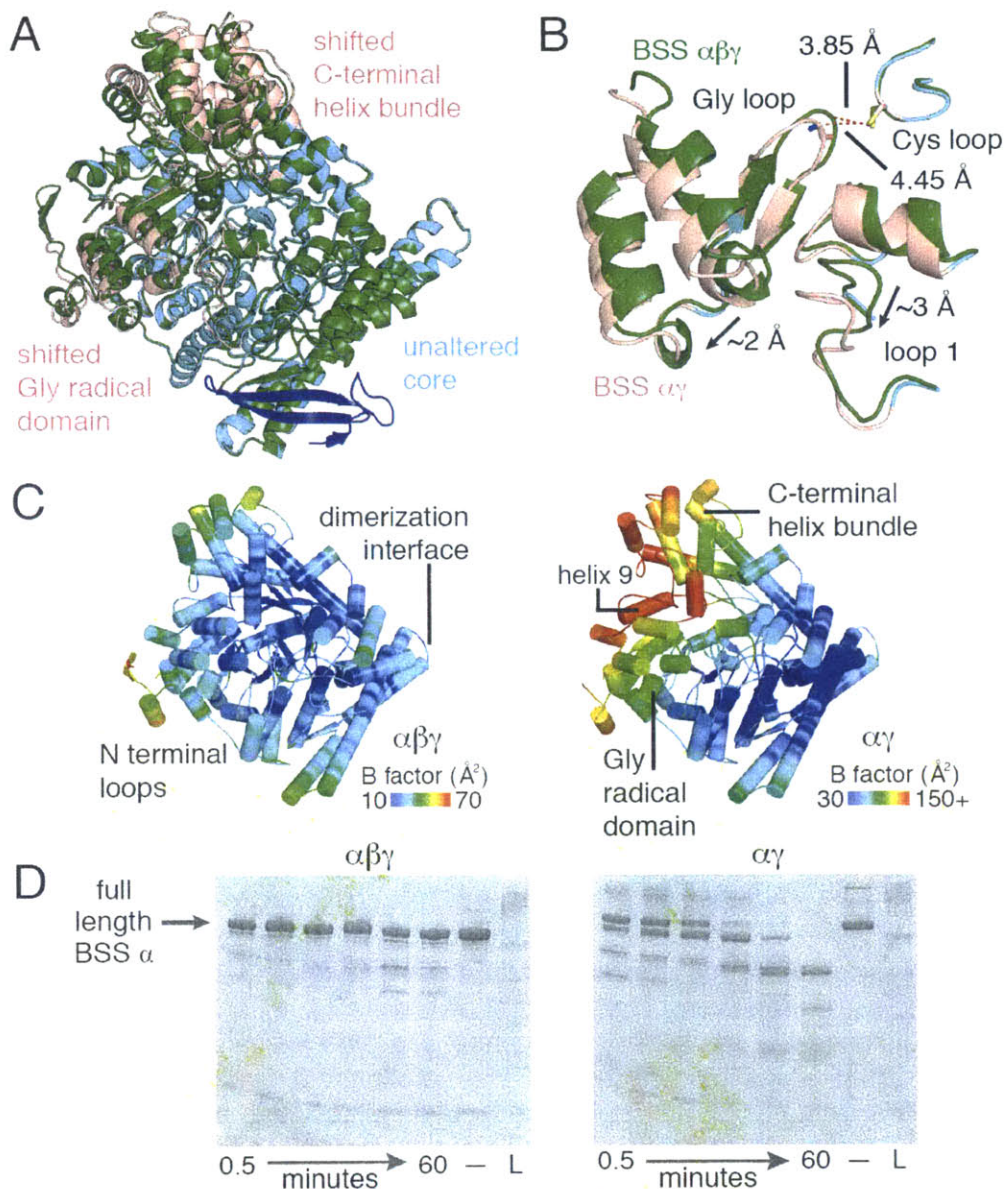


Figure II.12. Influence of BSS β on the structure and flexibility of BSS α . **A)** The α subunit from the $\alpha\beta\gamma$ complex (green) is overlaid with the $\alpha\gamma$ complex (blue or pink). The structures are aligned by the N-terminal half-barrel and dimerization helices, revealing conformational changes in the C-terminal half-barrel. BSS $\alpha\gamma$ Ca residues more than 1 Å from the corresponding Ca in BSS $\alpha\beta\gamma$ are colored pink, whereas those 1 Å or closer are colored blue. **B)** Cys loop, Gly loop, and loop 1 in the same color scheme as above, showing the increase in distance between the Cys S γ and Gly Ca from 3.85 Å to 4.45 Å due to absence of BSS γ . The average movement of the Gly radical domain and of loop 1 are indicated by arrows. **C)** B factors for the α subunit are shown according to the color scale. **D)** Anaerobic limited trypsin digestion of BSS complexes. Full-length BSS α is 98 kDa. The sample marked “—” was removed prior to addition of trypsin. The ladder is indicated by “L”.

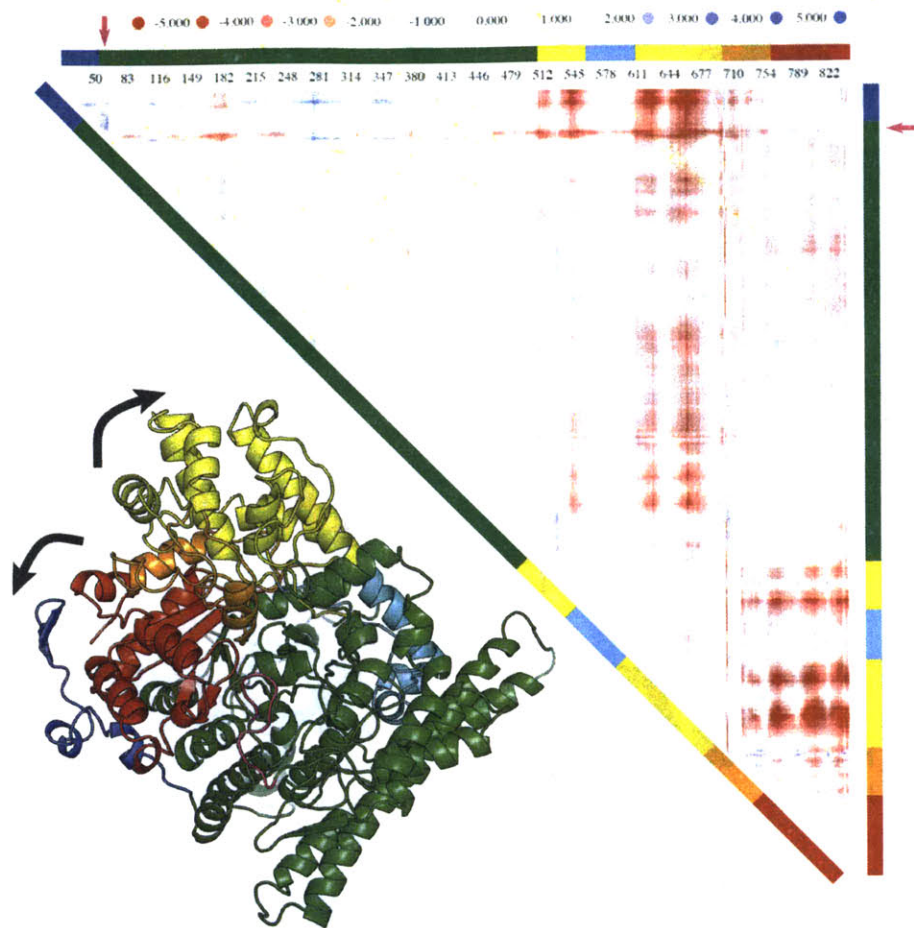


Figure II.13. Difference distance matrix showing differences in Ca distances between the BSS α β and BSS α complexes. The scale shows distance differences in Å such that red indicates closer distances in the BSS α β structure and blue indicates closer distances in the BSS α structure. White areas indicate regions with no change between Ca distances in the two structures. Ribbon drawing of BSS α shows the location of residues referred to in matrix, using consistent coloring. Regions that change between structures include the region from residue 510 to 700 that contains the C-terminal helix bundle (yellow) interrupted by part of helix 6 and strand 7 (teal) of the $(\beta\alpha)_{10}$ barrel. The C-terminal helix bundle moves but the residues of helix 6 and strand 7 are largely stationary. Residues 700-760, containing strand 8 and helix 9 (orange), are mostly disordered in BSS α and are less displaced than those of the C-terminal helix bundle. Loop 1 includes residues 59-68 (magenta loop in ribbon drawing and magenta arrow in matrix) and moves as a rigid body with the Gly radical domain. Residues 760-865 include strand 9 and 10, helix 10, and the Gly radical domain (red). A predominantly red plot indicates that all parts of the BSS α structure are moving away from each other, and that the structure is thus expanding. A lack of large blue areas indicates no contraction of the BSS α structure relative to the BSS α β structure. Many regions of the structure are stationary (white), for example, the N-terminal half-barrel. Adjacent clusters of red indicate large concerted motions of regions relative to each other. The arrows on the ribbon drawing indicate the clamshell motion described in the text. This plot was generated with the program DDMP from the Center for Structural Biology at Yale University.

The observed rearrangements described above prompted an analysis of the distribution of B factors in BSS α with and without BSS β present. Comparison of the two models clearly shows an increase in thermal motion around the C-terminal helix bundle and Gly radical domain in the BSS $\alpha\gamma$ complex (Figure II.12C). The N-terminal half-barrel and the long helices which form the dimer interface, in contrast, retain similar B factors and show hardly any conformational changes from the BSS $\alpha\beta\gamma$ structure. The highest B factors are along the center of the protein where the opening of the protein core destabilizes strand 9 and permits a range of motion to helix 9 and the surrounding loops.

Limited proteolysis of BSS complexes

Our crystallographic data indicate that BSS $\alpha\gamma$ is more open and flexible than BSS $\alpha\beta\gamma$, consistent with the hypothesis that BSS β regulates access to the active site in BSS α . An alternate explanation for the observed structural differences is that changes in the crystal lattice result in a greater degree of disorder in the BSS $\alpha\gamma$ structure. The BSS $\alpha\gamma$ crystal form does have fewer contacts to symmetry mates (four within 4 Å for $\alpha\gamma$, vs. nine within 4 Å for $\alpha\beta\gamma$) and shows a 10-20% increase in solvent content (39-43% in $\alpha\beta\gamma$ vs. 47% in $\alpha\gamma$) as well as an increase in average C α B factor (29.0 Å² for $\alpha\beta\gamma$ and 39.5 Å² for $\alpha\gamma$). To differentiate between the effects of BSS β and those of crystal packing on the flexibility of BSS α , we examined the flexibility of BSS α in solution by limited trypsin proteolysis. In agreement with the proposed role of BSS β , BSS $\alpha\gamma$ is degraded to numerous small fragments by trypsin in 1 h, whereas BSS $\alpha\beta\gamma$ is much more resistant to degradation over the same time period (Figure II.12D). Residues in BSS α are thus more accessible to trypsin when BSS β is absent. Although BSS β shields some regions in BSS α from proteolysis directly, the surface area covered by BSS β is only 1500 Å². This number is about 5% of the total BSS α surface area and not large enough to explain the almost complete degradation of BSS α in the absence of BSS β compared to almost none in its presence. The enhanced degradation of BSS α in the absence of BSS β therefore supports the conclusion that the BSS $\alpha\gamma$ complex is dynamic in solution as well as in our crystal structure.

II.D Discussion

The transformation of hydrocarbons into biologically-degradable compounds is chemically challenging, as these molecules lack functional groups that serve as handles for entry into

biological metabolism. In order to degrade inert substrates like toluene, microorganisms utilize enzymes capable of employing free radicals to functionalize these molecules for further oxidation. A fundamental question in the study of radical enzymes is how they control the generation and reactivity of radical species. Within the GRE family, structures have revealed a common mechanism for harnessing radicals: sequester them within a shell of protein residues that can exclude solvent and reactive small molecules, allowing for stable storage of the Gly radical for several days in vitro (26). For microbes that live both anaerobically and aerobically, burying the Gly radical within the protein core also provides some protection against oxygen-induced cleavage of the backbone at the radical site. This protection strategy, however, creates a problem for these enzymes: how to regulate access to the active site such that substrates are permitted in under the right conditions and other molecules are kept out. Furthermore, inactive GREs must undergo a large conformational rearrangement in order to interact with their cognate AE, but already activated enzymes must retain the Gly radical within the core of the protein to prevent radical quenching. This dichotomy of access and sequestration makes GREs an excellent target for studying the regulation of large conformational changes in enzymes.

Here we present structures of BSS complexes that illustrate a strategy for controlling conformational rearrangement in GREs: stabilization of a naturally flexible Gly radical domain by an accessory subunit. In the GRE structures so far determined, the Gly radical domain is consistently more flexible than the surrounding GRE core, as evidenced by a 5-10 Å² increase in the B factors for the domain relative to the average B factor for the remainder of the protein (Figure II.14). The BSSαγ complex recapitulates this relative increase in B factors in the Gly radical domain, suggesting that this complex is capable of accessing the open state and undergoing activation. In contrast, the BSSαβγ complex appears to be the outlier, having lower than average B factors overall and essentially no difference in B factors for the Gly radical domain relative to the rest of the protein. As seen in the BSSαγ structure, the Gly radical domain in BSSα appears more prone to enter the open state than in other GREs, and by binding to BSSβ, the resulting complex is stabilized in the catalytic state. Thus, BSSβ appears to modulate movement of the Gly radical domain, affecting both activation by the AE and protection of the stable Gly radical. Also, through its interactions with the Gly radical domain, the β subunit could regulate BSS activity by adjusting the distance between the Gly radical and catalytic Cys residue such that radical transfer is slowed, blocked or facilitated. Alternatively, or in addition, BSSβ

may regulate activity by gating entry of substrates to the active site through movement of the β hairpin loop.

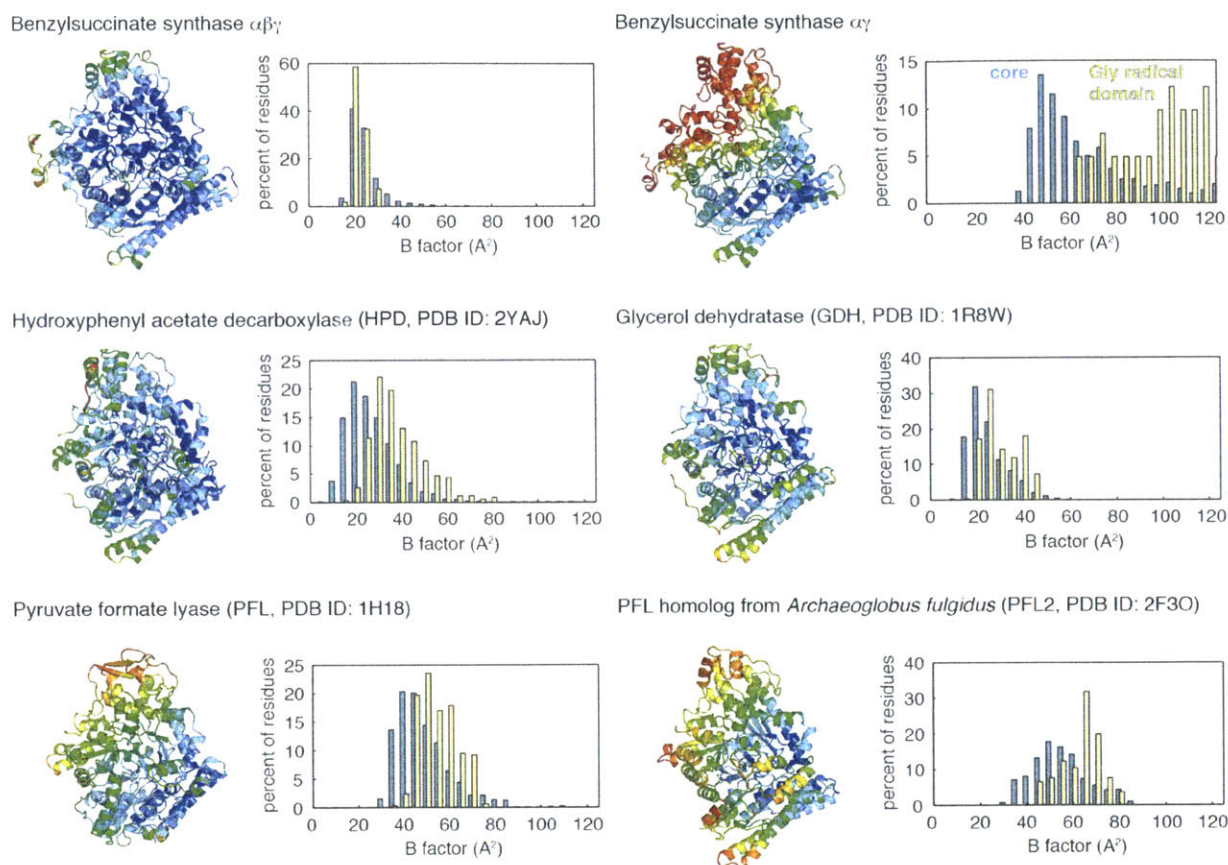


Figure II.14. B factor distribution in Gly radical enzymes. BSS α cartoon is colored by C α B factor from lowest (blue, $<10 \text{ \AA}^2$) to highest (red, $>100 \text{ \AA}^2$), with each structure scaled independently. The structures of hydroxyphenylacetate decarboxylase (19), pyruvate formate lyase (23), glycerol dehydratase (27), and a pyruvate formate lyase homolog (28) are representative structures from the GRE family. For each structure, a histogram of B factors of the main chain atoms of the GRE core (C, C α , N, O) is shown in blue, and the main chain of the Gly radical domain is shown in yellow. Where multiple molecules were present in the asymmetric unit, both monomers were included in the histogram.

The comparison of BSS with HPD deserves special attention because of the general similarity of these two enzymes: HPD is the closest relative to BSS with a solved structure; their substrates are similar in shape and properties; and they both have small, accessory, [4Fe-4S] cluster-containing subunits. Although the [4Fe-4S] cluster subunit of HPD, HpdC, has greater sequence similarity to BSS β , this subunit does not interact with the Gly radical domain, occupying instead a space near the dimer interface more reminiscent of how BSS γ binds to this

region of BSS α (Figure II.15A-B). Despite the differences in the location of their small subunits, BSS and HPD are similar in that they both have protein features absent in PFL and GDH that appear to prevent the Gly radical domain from easily interacting with the cognate AE (Figure II.16). In HPD, the Gly radical domain is blocked by an additional loop from the catalytic subunit, which overlaps with some of the contacts made by the β sheet of BSS β (Figure II.15A, II.16). This additional loop is not present in any other structurally characterized GREs and appears to be a short insertion into the pre-formed architecture of the GRE C-terminal helix bundle. Another loop that extends from a pair of helices near the dimer interface, which we will now call loop 2, is present in all structurally characterized GREs but is expanded in HPD to make extensive contact with the Gly radical domain (Figure II.16). In BSS, loop 1 and the C-terminus of BSS β make interactions with the Gly radical domain not seen in any other GRE (Figure II.15B, II.16). Loop 2 is present in BSS but makes no contacts to the Gly radical domain. Although these additional protein elements around the Gly radical domain in BSS and HPD are different, in both cases, they make the Gly radical domain considerably less solvent accessible than the other enzymes in their class (Figure II.16). HPD and BSS thus seem to have developed two different strategies for reducing the accessibility of the Gly radical loop. Notably, PFL, which appears to have the least encumbered radical domain (Figure II.16), has a repair mechanism to restore enzyme activity upon oxygen damage. In particular, repair involves a ‘spare part’ protein, YfiD, which is post-translationally activated by PFL-AE and replaces the oxygen-damaged glycy radical domain, yielding active enzyme (29). As more GRE enzymes are characterized, it will be interesting to see if structural features surrounding the glycy radical domain correlate with the likelihood of oxygen exposure or the presence of repair mechanisms.

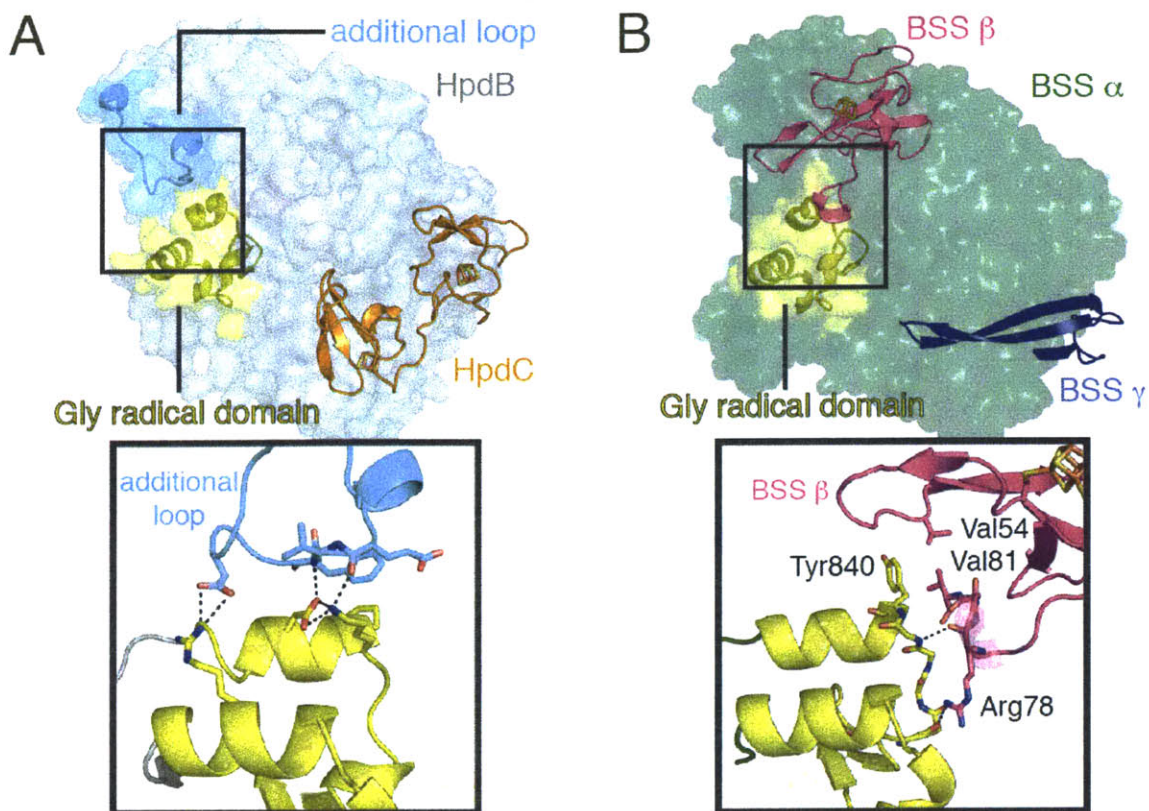
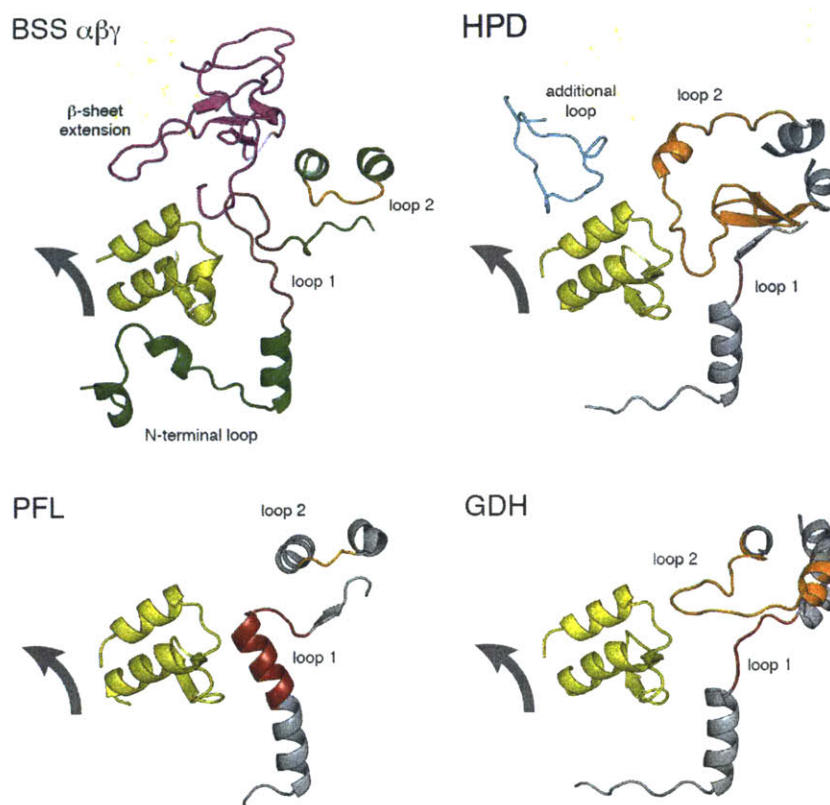


Figure II.15. Comparison of Gly radical domain interactions in BSS and HPD. **A)** The catalytic subunit of HPD (HpdB, gray) interacts with a small subunit (HpdC, orange) containing two [4Fe-4S] clusters. Instead of contacting HpdC, the Gly radical domain (yellow) interacts with a loop (cyan) that is not found in BSS. **B)** BSS α (green), shown with Gly radical domain in yellow, interacts with BSS β (magenta) and BSS γ (blue). BSS and HPD are shown in the same orientation. Insets: details of the Gly radical domain (yellow) for each structure. Tyr840, Val54 and Val81 in BSS make van der Waals contacts. Hydrogen bonds are shown as dashes.



	Gly radical domain, observed exposed surface area	surface area buried by additional structural elements			
		N-terminus	loop 1	loop 2/ BSS β C-terminus	additional loop/ BSS β β-sheet
BSS αβγ	1355	267	125	181	40
BSS αγ	1897	262	140		
HPD	1711			176	233
GDH	1924			100	
PFL	1782		185		

Figure II.16. Gly radical domain contacts across GRE family members. *Top*, interactions between Gly radical domains and neighboring loops are shown for four GRE enzymes: benzylsuccinate synthase, hydroxyphenylacetate decarboxylase (HPD) (19), pyruvate formate lyase (PFL) (23), and glycerol dehydratase (GDH) (27). The Gly radical domain was defined as follows: BSS: Val805-Glu850, HPD: Val849-Glu894, GDH: Ile739-Thr783, and PFL: Met710-Phe755. Loop 1 includes residues 55-65 of BSS α and residues 58-71 in PFL. Loop 1 does not contact the Gly radical domain in HPD or GDH. Loop 2 includes residues 98-113 in GDH, and the expanded version in HPD includes residues 128-169. Loop 2 does not contact the Gly radical domain in PFL or BSS. HPD has an additional loop (cyan) that none of these three other GRE enzymes have. Arrows show the proposed exiting motion of the Gly radical domain from the core of the protein. *Bottom*, table of surface areas for these four GRE enzymes. Surface areas (\AA^2) were calculated in PyMOL.

Although our structural data allows us to propose roles of the BSS subunits in regulating enzyme conformational dynamics and in modulating complex solubility, the particular function(s) of the [4Fe-4S] clusters in these subunits are less obvious. The clusters of both BSS β and γ do appear to play a structural role in that their removal leads to dissociation of the subunits from BSS α (17), which in turns leaves BSS α prone to unfolding and clearance by cellular degradation machinery *in vivo* (17), and sensitive to proteases *in vitro*. In terms of whether these clusters also play a redox role in addition to the above structural role, we find that the overall fold of both subunits is derived from a HiPIP, a family of [4Fe-4S] cluster proteins known for their high redox potentials (100-450 mV) and for their ability to stabilize clusters in the +3 oxidation state (21, 30). A universal pattern of hydrophobic shielding of the [4Fe-4S] clusters in HiPIPs is believed to be responsible for shifting the +3/+2 redox potential into a biologically relevant range, and shifting the +2/+1 couple out of a biological window, with potentials of -1.75 V (31). In contrast to ferredoxins, which utilize the +2/+1 couple, dithionite is typically unable to reduce HiPIP proteins (31). Here we find that BSS β and γ are hypermodified HiPIPs. Unlike most HiPIPs, the [4Fe-4S] clusters of BSS are not highly shielded from solvent, can be reduced by dithionite to the [4Fe-4S]¹⁺ states (17), and have redox couples that are proton-coupled. Like most HiPIP proteins, BSS β and γ are relatively stable in air and display high redox potentials, albeit on the lower end of the HiPIP range. In short, our data suggests that the [4Fe-4S] clusters of BSS can transition from oxidized to reduced to super-reduced ($[4\text{Fe-4S}]^{3+} \rightleftharpoons [4\text{Fe-4S}]^{2+} \rightleftharpoons [4\text{Fe-4S}]^{1+}$), an unusual, but not unprecedented, property for a protein with a HiPIP fold (31). These atypical redox properties support the consideration of redox roles as well as structural roles for the clusters of BSS.

For HPD, it has been proposed that the [4Fe-4S] clusters are redox active and could act to reductively quench the Gly radical by electron transfer, returning the enzyme to the inactive state (19). Efficient, reversible reduction of the Gly radical has been observed *in vitro* in PFL in the presence of DTT or β -mercaptoethanol (32), suggesting that Gly radical reduction could be a means of preventing permanent damage to the GRE by quenching the radical before O₂-catalyzed backbone cleavage can occur. However, no data are available to support the existence of this type of protective mechanism *in vivo*. In addition, the current structural data are inconsistent with such an electron transfer function, with long distances between the closest cluster and the Gly radical of over 40 Å in HPD and over 30 Å in BSS, whereas typical distances for electron transfer are in

the range of 8-12 Å (33). Notably, the distance between BSS β and the Gly of BSS α is likely to change when the radical domain undergoes its conformation change, potentially giving BSS β access to the Gly radical. Alternatively, the redox properties of the [4Fe-4S] clusters could be used to sense the redox state of the cell and communicate to BSS α , either by initiating a conformational change or altering the solubility of BSS α .

In summary, although movement of the Gly radical domain out of the core of the protein was previously proposed (14), our structures of BSS provide the first experimental evidence that this domain can move as a unit from the protein interior toward the protein surface. Prior to these structures, it was not known whether other adjustments to the protein core would be required for this motion of the Gly radical domain. We now see that the C-terminal half-barrel exhibits a 5-7 Å clamshell-like motion centered at strand and helix 9 to expand the core enough for the Gly radical domain to egress. These structures also allow us to present several new hypotheses as to the function of these [4Fe-4S] cluster-containing accessory subunits that include modulation of conformational dynamics and structural stability as a means of regulating enzyme activation and activity. Given the potential applications of GREs in environmentally and industrially important transformations of inert hydrocarbons, these findings provide a valuable framework for considering the protein design features that harness radical-based reactivity in this significant class of anaerobic enzymes.

II.E Materials and Methods

Protein preparation

C-terminally His-tagged BSS α was co-expressed with either BSS γ alone, or BSS γ and BSS β and purified as previously described (17). For crystallography experiments, a single Ni-nitriloacetic acid affinity chromatography step was used to rapidly purify the protein aerobically. All column buffers were made anaerobic by bubbling with argon prior to use, and 0.5 mM *tris*(2-carboxyethyl)phosphine (TCEP) was included in all steps to maintain a reducing environment. Cluster content of the small subunits was monitored by absorbance at 410 nm at the end of the purification and was comparable to proteins purified anaerobically. Protein for electrochemistry and degradation assays was purified anaerobically. Selenomethionine (SeMet) protein for phasing was produced in minimal medium using a metabolic inhibition scheme previously described by our laboratory (34).

Crystallization

Crystals of BSS $\alpha\beta\gamma$ were grown by sitting drop vapor diffusion at 25 °C in a Coy anaerobic chamber under an atmosphere of 95% Ar and 5% H₂. Protein at ~8 mg/mL in buffer containing 50 mM Tris pH 7.6, 15% (vol/vol) glycerol, 200 mM NaCl, and 0.03% *N*-lauroylsarcosine was added at a 2:1 ratio to well solution containing 25% (wt/vol) PEG 3350, 100 mM Tris pH 8.5, and 200 mM NH₄(CH₃CO₂). Diffraction-quality crystals typically appeared after 1-2 weeks in optimization trays. SeMet crystals were grown similarly, but with 17-20% (wt/vol) PEG 3350, 100 mM KCl, and 50 mM Tris pH 8.5 as the well solution. Crystals were soaked for 2 min in 10% (vol/vol) glycerol, 25% (wt/vol) PEG 3350, 50 mM Tris pH 8.5 for cryoprotection and plunged in liquid nitrogen while in the anaerobic chamber.

Crystals of BSS $\alpha\gamma$ were grown by hanging drop vapor diffusion at 20 °C in an MBraun anaerobic chamber with O₂ maintained below 0.1 ppm. Protein at ~10 mg/mL in buffer containing 20 mM HEPES pH 7.5 and 100 mM NaCl was incubated at a 1:1 ratio with well solution containing 20-25% (wt/vol) PEG 400, 50 mM Tris pH 8.0, and 50 mM Bis-Tris pH 6.5. Diffraction quality crystals typically grew within 1 week. Crystals were briefly soaked in cryoprotectant solution containing 30% (wt/vol) PEG 400 and the same buffer.

Data processing and phasing

Statistics for each dataset collected are given in Table 1. Crystals of BSS $\alpha\beta\gamma$ indexed as either I222 ($a = 113.4$, $b = 120.4$, $c = 136.0$ Å) with one molecule in the asymmetric unit or P2₁2₁2 ($a = 139.8$, $b = 116.2$, $c = 124.2$ Å) with two molecules in the asymmetric unit. Later analysis revealed the two crystal forms have virtually identical packing with small differences that disrupt the translational symmetry of the body-centered form to give the primitive form. Indexing, integration, and scaling were performed in HKL2000 (35). The structure of BSS $\alpha\beta\gamma$ was solved by the single wavelength anomalous dispersion technique (SAD) with data collected at the selenium peak wavelength, 0.9792 Å. A single dataset was collected on an ASDC Quantum-4 detector in inverse beam mode with 10° wedges. Anomalous signal extended to 3.6 Å and all initial phasing steps were performed at that resolution. Initial Se sites were found with HKL2MAP (36) and input into Phenix AutoSol (37) for site refinement, phasing, iterative density modification, and phase extension to the full extent of the data. Experimental maps with density modification at 3.1-Å resolution were sufficient to manually build the entire protein model in COOT (38). The completed model was transferred to the high-resolution native data by molecular replacement in Phenix AutoMR (37).

BSS $\alpha\gamma$ crystallized in space group P4₃2₁2 ($a = 154.9$, $b = 154.9$, $c = 82.2$ Å) with a single molecule in the asymmetric unit. Data were collected on a Quantum-4 ASDC CCD detector at 1.000 Å wavelength. The scaled data were assessed to be strongly anisotropic by the Diffraction Anisotropy Server (<http://services.mbi.ucla.edu/anisoscale/>), and were thus truncated and rescaled to compensate (39). The resulting maps were markedly improved. The structure of BSS $\alpha\gamma$ was solved by molecular replacement with the program Phaser (40) using only the α and γ chains of the BSS $\alpha\beta\gamma$ structure and a 3.5-Å resolution cutoff. A similar dimer interface is generated by symmetry in both $\alpha\beta\gamma$ and $\alpha\gamma$ structures, consistent with the dimer form of the enzyme being stable and relevant.

Refinement

For the BSS $\alpha\beta\gamma$ structure, individual B factor, translation/libration/screw, and positional refinement was performed with phenix.refine (37), and model building and addition of water molecules was performed in COOT. A final round of refinement was performed in CNS (41) and gave similar statistics with insignificant changes to the model. CNS simulated annealing

composite omit maps were used to verify side chain positions and the occupancy of ligands and water molecules. In BSS α , 8 N-terminal residues and 15 C-terminal residues, including the C-terminal His₆-tag, were missing in the final model. BSS β and - γ were missing 15 and 10 N-terminal residues and 0 and 12 C-terminal residues, for a model completeness of 81 and 63%, respectively. The final BSS $\alpha\beta\gamma$ model contained 97.3% of residues (933 residues) in the most favored region of the Ramachandran plot, with 2.4% additionally allowed (22 residues) and 0.3% outliers (3 residues). Ramachandran outliers were examined with composite omit maps and appear to be correctly refined. The Cys S γ bond distances to the [4Fe-4S] cluster in BSS β were restrained to a distance of 2.3 Å with $\sigma=0.1$ Å.

The BSS $\alpha\gamma$ structure was refined initially in CNS v1.3 using deformable elastic network (DEN) restraints (42), which allow large structural rearrangement while maintaining the overall fold of a protein domain. Refinement was completed with positional and B factor refinement in phenix.refine with manual model building in COOT. Composite omit maps reveal that residues 706-715, 731-765, and 840-865 are largely disordered. Omission of these residues causes difference density to appear in their place and an increase in the overall R_{free} , indicating that they should be retained. Reference model restraints were used to limit their divergence from the observed structure in the BSS $\alpha\beta\gamma$ structure. The B factors for these residues are unusually high, reflecting their poor ordering in the structure. Those residues missing from the N and C termini of BSS α and - γ in the BSS $\alpha\beta\gamma$ structure are likewise missing in the BSS $\alpha\gamma$ structure, along with 10 additional residues at the N terminus of BSS α . The final BSS $\alpha\gamma$ model contained 96.4% of residues in the most favored region of the Ramachandran plot, with 3.4% additionally allowed and 0.2% outliers (3 residues). Ramachandran outliers are in disordered regions of the protein and could not be improved with additional refinement.

Table 1. Data collection and refinement statistics.

	SeMet BSS $\alpha\beta\gamma$	BSS $\alpha\beta\gamma$	BSS $\alpha\gamma$
Data collection and processing			
Space group	P2 ₁ 2 ₁ 2	I222	P4 ₃ 2 ₁ 2
Cell dimensions (Å)	139.8, 116.2, 124.2	113.4, 120.4, 136.0	154.9, 154.9, 82.2
Beamline	APS ¹ 24-ID-C	ALS ² 8.2.2	NSLS ³ X26C
Dataset	SAD peak	native	native
Wavelength (Å)	0.9792 (Se K edge)	1.0000	1.0000
Resolution (Å)	50.0–3.10 (3.21–3.10)	50.0-2.00 (2.03-2.00)	49.0-2.60 (2.64-2.60)***
Completeness (%)	99.9 (99.9)*	99.2 (87.7)**	98.1 (85.1)***
$\langle I/\sigma \rangle$	17.1 (3.2)*	21.4 (9.6)	10.7 (1.1)***
R _{sym}	0.12 (0.56)*	0.084 (0.29)	0.15 (>1.0)***
Unique reflections	75353*	62251	30845
Redundancy	8.1 (7.9)*	14.3 (11.0)	7.0 (4.8)
CC1/2		(0.972)	(0.606)
Phasing			
Phasing resolution ⁴	43.95–3.60 (3.69–3.60)		
Figure of merit (FOM)	0.54 (0.375)		
Phasing power	2.65 (1.10)		
Density modification FOM ⁴	0.74 (0.55)		
Model refinement			
Resolution (Å)		49.5-2.00	49.0-2.60
R _{work} , R _{free} ⁵		0.129, 0.177	0.217, 0.257
Protein atoms		7717	6992
Water molecules		836	194
Bond length deviation (Å)		0.013	0.002
Bond angle deviation (°)		1.318	0.57
Rotamer outliers (%)		0.4	0.0
Average B factors (Å ²)			
BSS α		28.1	85.6
BSS β		38.1	
BSS γ		38.0	74.3
water		36.6	58.1

¹Advanced Photon Source, Argonne National Laboratory

²Advanced Light Source, Lawrence Berkeley National Laboratory

³National Synchrotron Light Source, Brookhaven National Laboratory

⁴Resolution was extended to 3.10 Å in final rounds of density modification.

⁵3000 reflections were set aside for cross validation.

*Statistics were calculated with Friedel pairs not merged.

**Data were truncated at 2.0-Å resolution due to low completeness beyond this point.

***Data were strongly anisotropic along the c axis, resulting in poor statistics overall. The quality of the data was verified by the Diffraction Anisotropy Server.

Limited proteolysis

As isolated BSS $\alpha\beta\gamma$ and BSS $\alpha\gamma$ (2 $\mu\text{g}/\mu\text{L}$) were incubated in buffer containing 20 mM Tris pH 8.0, 1.25 mM CaCl_2 , 0.5 mM TCEP and 5 ng/ μL bovine pancreatic trypsin (Sigma) at 15 °C in an MBraun anaerobic chamber with $\text{O}_2 < 0.1$ ppm. Samples were removed at time points ranging from 30 s to 1 h and quenched by addition of N α -tosyl-L-lysyl chloromethyl ketone to 30- μM final concentration. Sodium dodecyl sulfate loading dye was added and samples were denatured at 90 °C and run on "Any kD" Mini-PROTEAN TGX gels (BioRad) to visualize the fragment bands.

Electrochemistry

Electrochemistry experiments were performed using a PGSTAT-12 or PGSTAT-30 potentiostat (Ecochemie) equipped with FRA and ECD modules. Data were collected both aerobically and in an anaerobic chamber (MBraun) at pH 7, 20 °C, 60 mV/s. A three-electrode electrochemistry cell with a platinum wire as a counter electrode and a calomel reference electrode were used. Experiments were performed in a mixed buffer containing 5 mM each [sodium acetate, 2-(N-morpholino)ethanesulfonic acid, 3-morpholinopropane-1-sulfonic acid, 3-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]propane-1-sulfonic acid, 2-(cyclohexylamino)ethanesulfonic acid, 3-(cyclohexylamino)-1-propanesulfonic acid] and 100 mM NaCl (pH 7). For pH dependence experiments 10 mM rather than 5 mM buffering agents were used. Enzyme films were generated by depositing 3 μL enzyme solution onto an alkanethiol-modified gold electrode, waiting 20-30 s, then removing excess protein solution. Analysis of electrochemical data was performed using the SOAS software package (43). Gold electrodes were prepared by polishing with 3- μm crystalline diamond suspension followed by 1, 0.5, and 0.03 μm alumina. After a brief sonication electrodes were electrochemically cleaned by cycling between 0.2 and 1.35 V (v SCE) in 0.1 M H_2SO_4 . Alkanethiol-modified gold surfaces were generated by soaking electrodes overnight in solutions of 55 mM 3-mercaptopropionic acid (3-MPA) in ethanol. Electrodes were then rinsed with water and ethanol, and enzyme films were generated.

II.F Acknowledgements

This work was supported in part by the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. 0645960 (M.A.F.), NSF Grant MCB-0543833 (C.L.D.), National Institutes of Health (NIH) GM093088 (E.N.G.M), and NIH Research Service Award GM099416 (E.T.J.). C.L.D is a Howard Hughes Medical Institute (HHMI) Investigator. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by award NIH GM103403 from the National Center for Research Resources at NIH. Use of the Advanced Photon Source is supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences (BES), under Contract No. DE-AC02-06CH11357. Portions of this work were performed at Beamline X26A, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. X26A is supported by DOE Geosciences (DE-FG02-92ER14244 to The University of Chicago CARS). Use of the NSLS was supported by DOE BES, under Contract No. DE-AC02-98CH10886. Portions of this work were also carried out at the Berkeley Center for Structural Biology, which is supported in part by the NIH GM, and HHMI. The Advanced Light Source is supported by U.S. DOE BES under Contract No. DE-AC02-05CH11231.

II.G References

1. Fuchs G (2008) Anaerobic metabolism of aromatic compounds. *Ann N Y Acad Sci* 1125:82-99.
2. Biegert T, Fuchs G, & Heider J (1996) Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem* 238(3):661-668.
3. Beller HR & Spormann AM (1998) Analysis of the novel benzylsuccinate synthase reaction for anaerobic toluene activation based on structural studies of the product. *J Bacteriol* 180(20):5454-5457.
4. Qiao C & Marsh EN (2005) Mechanism of benzylsuccinate synthase: stereochemistry of toluene addition to fumarate and maleate. *J Am Chem Soc* 127(24):8608-8609.
5. Li L & Marsh EN (2006) Mechanism of benzylsuccinate synthase probed by substrate and isotope exchange. *J Am Chem Soc* 128(50):16056-16057.
6. Li L & Marsh EN (2006) Deuterium isotope effects in the unusual addition of toluene to fumarate catalyzed by benzylsuccinate synthase. *Biochemistry* 45(46):13932-13938.

7. Krieger CJ, Roseboom W, Albracht SP, & Spormann AM (2001) A stable organic free radical in anaerobic benzylsuccinate synthase of *Azoarcus* sp. strain T. *J Biol Chem* 276(16):12924-12927.
8. Widdel F & Rabus R (2001) Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Curr Opin Biotechnol* 12(3):259-276.
9. Acosta-Gonzalez A, Rossello-Mora R, & Marques S (2013) Diversity of benzylsuccinate synthase-like (bssA) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* 79(12):3667-3676.
10. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.
11. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.
12. Knappe J, Neugebauer FA, Blaschkowski HP, & Ganzler M (1984) Post-translational activation introduces a free radical into pyruvate formate-lyase. *Proc Natl Acad Sci USA* 81(5):1332-1335.
13. Shisler KA & Broderick JB (2014) Glycyl radical activating enzymes: Structure, mechanism, and substrate interactions. *Archives of Biochemistry and Biophysics* 546(0):64-71.
14. Vey JL, *et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.
15. Peng Y, Veneziano SE, Gillispie GD, & Broderick JB (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J Biol Chem* 285(35):27224-27231.
16. Leuthner B, *et al.* (1998) Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* 28(3):615-628.
17. Li L, *et al.* (2009) Subunit structure of benzylsuccinate synthase. *Biochemistry* 48(6):1284-1292.
18. Hilberg M, *et al.* (2011) Identification of FeS clusters in the glycyl-radical enzyme benzylsuccinate synthase via EPR and Mössbauer spectroscopy. *J Biol Inorg Chem* 17(1):49-56.

19. Martins BM, *et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *J Am Chem Soc* 133(37):14666-14674.
20. Coschigano PW (2002) Construction and characterization of insertion/deletion mutations of the *tutF*, *tutD*, and *tutG* genes of *Thauera aromatica* strain T1. *FEMS Microbiol Lett* 217(1):37-42.
21. Heering HA, Bultink BM, Hagen WR, & Meyer TE (1995) Influence of charge and polarity on the redox potentials of high-potential iron-sulfur proteins: evidence for the existence of two groups. *Biochemistry* 34(45):14675-14686.
22. Tilley GJ, Camba R, Burgess BK, & Armstrong FA (2001) Influence of electrochemical properties in determining the sensitivity of [4Fe-4S] clusters in proteins to oxidative damage. *Biochem J* 360(Pt 3):717-726.
23. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA. How the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.
24. Chovancova E, *et al.* (2012) CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* 8(10):e1002708.
25. Richards FM & Kundrot CE (1988) Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure. *Proteins: Structure, Function, and Bioinformatics* 3(2):71-84.
26. Walsby CJ, *et al.* (2005) Spectroscopic approaches to elucidating novel iron-sulfur chemistry in the "radical-Sam" protein superfamily. *Inorg Chem* 44(4):727-741.
27. O'Brien JR, *et al.* (2004) Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.
28. Lehtio L, Grossmann JG, Kokona B, Fairman R, & Goldman A (2006) Crystal structure of a glycyl radical enzyme from *Archaeoglobus fulgidus*. *J Mol Biol* 357(1):221-235.
29. Wagner AF, *et al.* (2001) YfiD of *Escherichia coli* and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem Biophys Res Commun* 285(2):456-462.
30. Perrin BS, Jr & Ichiye T (2010) Fold versus sequence effects on the driving force for protein-mediated electron transfer. *Proteins* 78(13):2798-2808.

31. Heering HA, Bulsink YB, Hagen WR, & Meyer TE (1995) Reversible super-reduction of the cubane [4Fe-4S](3+;2+;1+) in the high-potential iron-sulfur protein under non-denaturing conditions. EPR spectroscopic and electrochemical studies. *Eur J Biochem* 232(3):811-817.
32. Nnyepi MR, Peng Y, & Broderick JB (2007) Inactivation of *E. coli* pyruvate formate-lyase: role of AdhE and small molecules. *Arch Biochem Biophys* 459(1):1-9.
33. Moser CC, Keske JM, Warncke K, Farid RS, & Dutton PL (1992) Nature of biological electron transfer. *Nature* 355(6363):796-802.
34. Ryan KS, *et al.* (2007) Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC. *Proc Natl Acad Sci U S A* 104(39):15311-15316.
35. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data. *Methods in Enzymology*, Vol 276.
36. Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* 37(5):843-844.
37. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
38. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
39. Strong M, *et al.* (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 103(21):8060-8065.
40. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
41. Brunger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5):905-921.
42. Schroder GF, Levitt M, & Brunger AT (2010) Super-resolution biomolecular crystallography with low-resolution data. *Nature* 464(7292):1218-1222.
43. Fourmond V, *et al.* (2009) SOAS: a free program to analyze electrochemical data and other one-dimensional signals. *Bioelectrochemistry* 76(1-2):141-147.

Chapter III

The structure of a pre-turnover complex of benzylsuccinate synthase guides mechanistic understanding of the X-succinate synthase enzyme family

A similar manuscript has been submitted to the *Proceedings of the National Academy of Sciences*. This chapter was written with Prof. Catherine L. Drennan. Prof. E. Neil G. Marsh (University of Michigan) provided constructs and reagents.

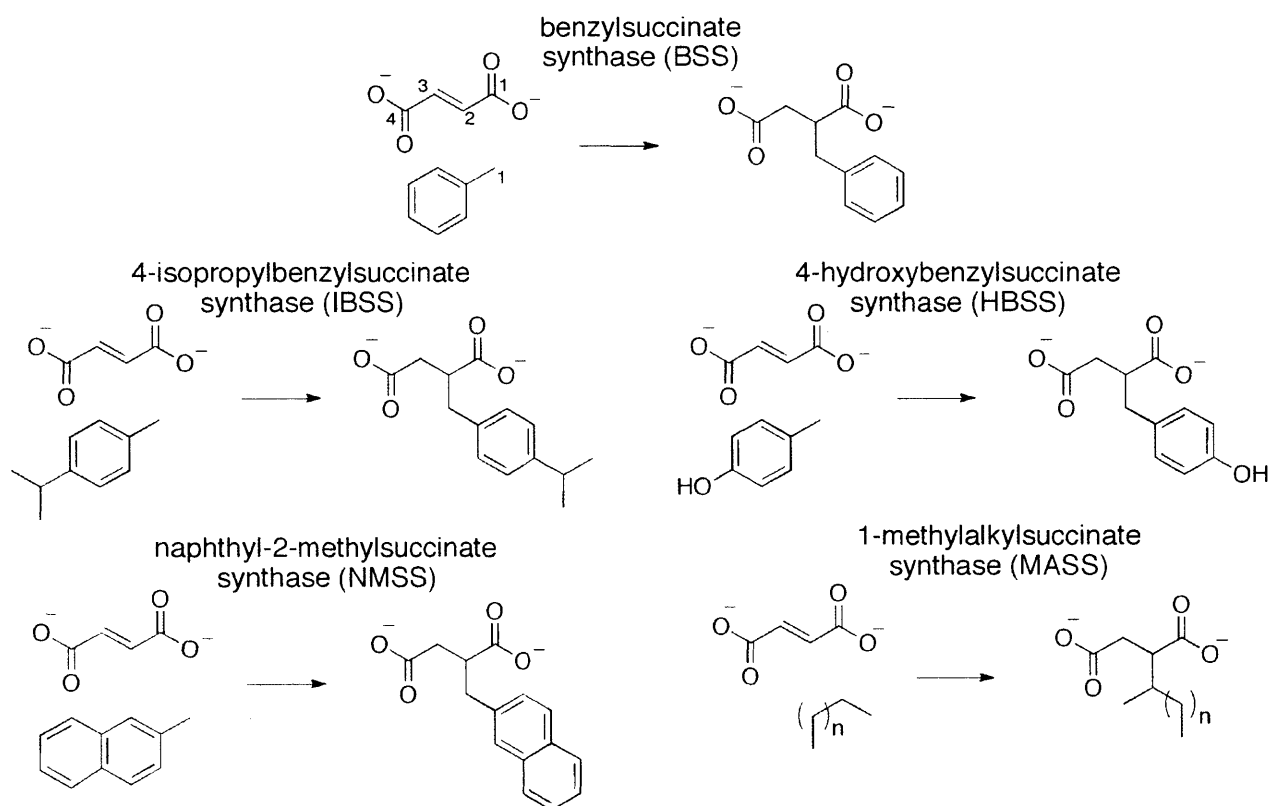
III.A Summary

Bacteria perform anaerobic degradation of a variety of small hydrocarbons as a source of energy and cellular carbon. Anaerobic toluene degradation involves conjugation of toluene to a molecule of fumarate in a radical-catalyzed, C–C bond-forming reaction. We have determined structures of the heterohexameric, $(\alpha\beta\gamma)_2$ enzyme that performs this reaction, benzylsuccinate synthase (BSS), in two oligomeric states with fumarate or both fumarate and toluene bound. We find that the presence of the small subunit BSS β is necessary for both fumarate and toluene to bind stably to BSS α , the catalytic subunit. BSS β inserts into a deep channel that leads to the active site and appears to regulate substrate binding and stabilize the catalytically closed form of the enzyme. Fumarate is secured at the bottom of the active site through hydrogen-bonding interactions and packing of the double bond between two active site residues. Toluene binds within a deeply buried cavity in an orientation set up to facilitate C–C bond formation during the radical-based mechanism. A radical storage residue—a backbone glycyl radical—lines up perfectly for hydrogen atom transfer with an active site cysteine, and the resulting thiyl radical abstracts a hydrogen from toluene to initiate the reaction. Tight packing of hydrophobic residues around toluene ensures hydrocarbon substrate specificity in BSS. Enzymes related to BSS are able to metabolize a wide range of hydrocarbons through attachment to fumarate. Using our structures as a guide, we have constructed homology models of several of these enzymes and determined conservation patterns that will be useful in understanding the enzymatic basis for catalysis and specificity in the broader X-succinate synthase family.

III.B Introduction

Microbial degradation is one route for the mitigation of hydrocarbon pollution (1-3). Bacterial species have devised numerous means of degrading inert hydrocarbon molecules (4). In the presence of oxygen, a wide range of bacteria, and some fungi, initiate metabolism of aromatic hydrocarbons and alkanes through direct oxidation by mono- or dioxygenases, followed by β oxidation of the resulting hydroxylated fragments. In the absence of oxygen, hydrocarbon functionalization is more challenging, and different strategies are required. Anaerobic degradation of toluene by the denitrifying bacterium *Thauera aromatica* is the best-studied microbial hydrocarbon degradation pathway thus far. The first step in this pathway is conversion of toluene into *R*-benzylsuccinate, catalyzed by the enzyme benzylsuccinate synthase (BSS),

with fumarate as a co-substrate (5). Benzylsuccinate, having a carboxylate functional handle, can be degraded through β oxidation to benzoyl-CoA and then reductively dearomatized by benzoyl-CoA reductase prior to further oxidation (4, 6). A group of enzymes closely related to BSS, collectively dubbed aryl-succinate synthase (SS) enzymes, are known to function on toluene-like aromatic hydrocarbons, including *p*-cymene (7, 8), *p*-cresol (9, 10), and 2-methylnaphthalene (11, 12) (Scheme III.1). Additionally, alkyl-SSs, which function on alkyl chains including *n*-hexane, have also been discovered (13, 14), although there is less known about the substrate specificity and mechanism of these enzymes. Together these enzymes form a large family of X-succinate synthases (XSSs) responsible for functionalization of a wide variety of hydrocarbon substrates, and the range of substrates is still growing.



Scheme III.1. Reactions catalyzed by characterized X-succinate synthases (XSSs). Most characterized XSSs are aryl-SSs (BSS, IBSS, HBSS, NMSS), but non-aryl family members have also been discovered (MASS).

The amazing transformation performed by BSS and its relatives, which is a unique C–C bond-forming reaction in biology, requires an oxygen-sensitive radical cofactor. BSS is a

member of the glycy radical enzyme (GRE) family and contains a backbone glycy radical in its activated form. Other members of this family include the central metabolic enzymes pyruvate formate-lyase (PFL) and anaerobic ribonucleotide reductase (class III or anRNR) and several radical eliminases (15, 16). All of these radical enzymes are believed to generate a substrate-based radical that undergoes rearrangement in the active site to generate a product-based radical, which likewise regenerates the glycy radical. In all GREs thus far, an enzyme-based thiyl radical is proposed to be an intermediary between the glycy- and substrate-based radical species (17, 18). The cysteine proposed to form the thiyl radical initiates radical chemistry on the substrate, whereas the glycine is the radical storage site.

Radical chemistry in GREs occurs deep within the core of a 10-stranded β/α barrel where reactive radical species are protected. At the barrel core, a loop containing the essential glycine—the Gly loop—is juxtaposed with the loop containing the catalytic cysteine—the Cys loop—allowing for hydrogen atom transfer between the two. Residues surrounding these loops comprise the active site. BSS exists as an $(\alpha\beta\gamma)_2$ heterohexamer (Figure III.1). BSS α contains the 10-stranded barrel and is the catalytic subunit, whereas BSS β and BSS γ are both small, [4Fe-4S]-cluster-containing proteins of unknown function (19, 20). Despite the fact that the roles of BSS β and BSS γ are unclear, both proteins are required for viable growth of *Thauera aromatica* on toluene (21, 22). We previously solved structures of the BSS $\alpha\beta\gamma$ and BSS $\alpha\gamma$ complexes, revealing that removal of BSS β from the complex with BSS $\alpha\gamma$ permits a large conformational change within BSS α (23). The core barrel opens, exposing the active site, and a C-terminal domain that harbors the Gly loop—the glycy radical domain—shifts a few ångströms out of the active site but does not completely vacate the barrel. Movement of the glycy radical domain away from the barrel core is expected to be required for the initial installation of the glycy radical by the cognate AdoMet radical activating enzyme (24). Thus, in the absence of BSS β , the enzyme adopts an open state that appears amenable to activation, whereas in the presence of BSS β , the enzyme adopts a closed state that appears ready for catalysis, with the glycy radical domain positioned for radical transfer. We identified a channel that substrates could take to enter the active site; however, BSS β appeared to block the channel entrance, suggesting that either substrates cannot bind in the presence of BSS β or that BSS β plays a role in gating active site accessibility.

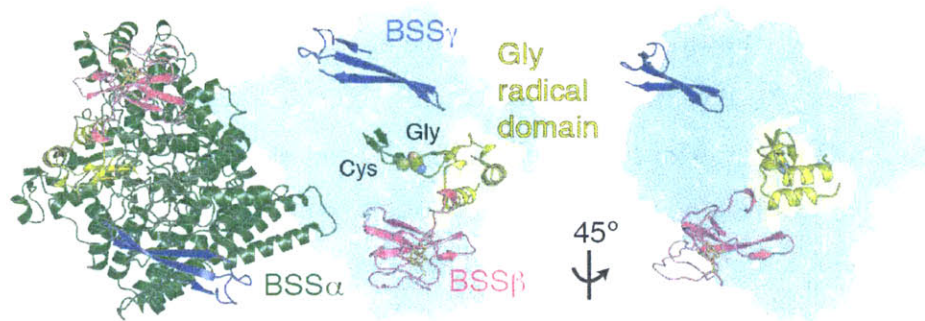


Figure III.1. BSS forms an $(\alpha\beta\gamma)_2$ heterohexamer. BSS α , the large, catalytic subunit forms a dimer through a large interface. BSS β and BSS γ , the small, [4Fe-4S]-cluster-containing subunits, bind at independent sites within a single BSS α protomer. In this structure only BSS β contains a cluster due to close crystal contacts. The Cys loop and Gly loop are shown at the center of the surface view and are deeply buried within the core barrel fold of the enzyme. The Gly loop resides on the glycy radical domain, which inserts into the barrel on one face. BSS β rests on top of the glycy radical domain and makes additional contacts through a short C-terminal tail. *Figure adapted from (23).*

In this chapter, we investigate how BSS binds to fumarate and toluene in both $\alpha\beta\gamma$ and $\alpha\gamma$ complexes. We find that both substrates can bind to the BSS $\alpha\beta\gamma$ complex, thus demonstrating that the presence of BSS β does not completely prevent substrate access into the active site. Fumarate binding to BSS $\alpha\gamma$ partially shifts the barrel from the open state observed in the substrate-free BSS $\alpha\gamma$ structure toward the closed state seen in BSS $\alpha\beta\gamma$. However, ordering of BSS α into the fully closed, catalytic state does not occur in this structure in the absence of BSS β ; only structures with BSS $\alpha\beta\gamma$ depict the fully closed, catalytically competent state. In this state, both substrates are bound at the bottom of the proposed channel in an orientation consistent with hydrogen atom abstraction from toluene by the putative, transient thiyl radical, followed by C–C bond formation between toluene and fumarate. Finally, we have constructed homology models of other members of the XSS family based on the structure of BSS in order to reveal the determinants of specificity in these enzymes. This analysis will aid in characterization of the diverse communities of microbes known to cooperate in the degradation of hydrocarbons.

III.B Results

Substrates can bind to BSS $\alpha\beta\gamma$.

To investigate how substrates bind, we incubated BSS $\alpha\beta\gamma$ with fumarate and toluene prior to crystallization. Crystals produced by this method diffract to 3.3-Å resolution. Omit

electron density peaks in the central active site can be modeled as molecules of fumarate and toluene (Figure III.2A). Substrates-bound BSS $\alpha\beta\gamma$ adopts the same closed state observed previously in substrate-free BSS $\alpha\beta\gamma$ (23) with virtually no structural changes observable within the active site or on the exterior of the protein (BSS α C α root-mean-square deviation: 0.057 Å). As expected, the substrates are bound at the bottom of the proposed access channel. This channel is formed by two helices outside the core barrel, $\alpha 3'$ and $\alpha 8'$, and a loop between $\alpha 1'$ and $\beta 1$. Together these elements contribute residues V185, E189, A308, Y381, I384, and L711 (Figure III.3), forming a contraction at the midway point that appears to restrict movement of substrates and solvent in or out of the active site. In addition to residues from BSS α , a hairpin loop from BSS β is inserted within the putative access channel and appears to completely block access to the active site. Although these structural data show that substrates can bind to BSS in the presence of BSS β , it remains unclear what conformational changes are required for passage of substrates into the active site within the full BSS $\alpha\beta\gamma$ complex.

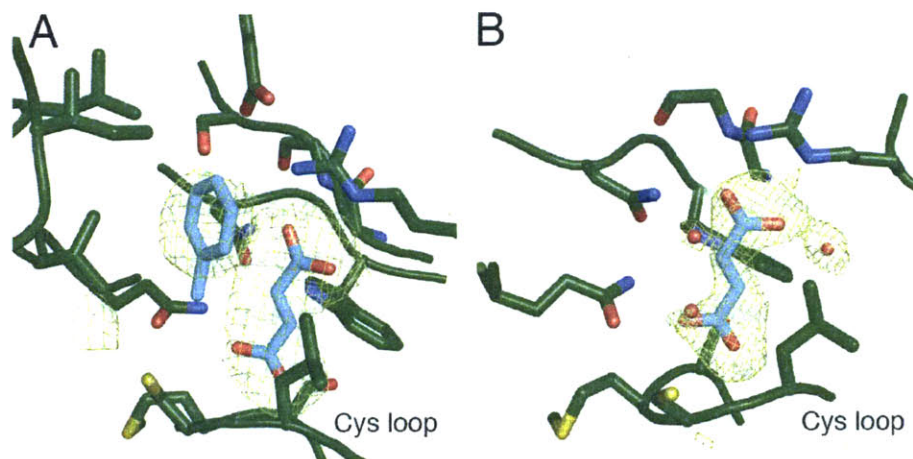


Figure III.2. Electron density peaks from bound substrates are present in BSS structures. (A) Fumarate and toluene co-crystallization in BSS $\alpha\beta\gamma$. (B) Fumarate co-crystallization in BSS $\alpha\gamma$. mF_o-DF_c omit density is shown at 3σ from the fully refined structures after exclusion of the ligands.

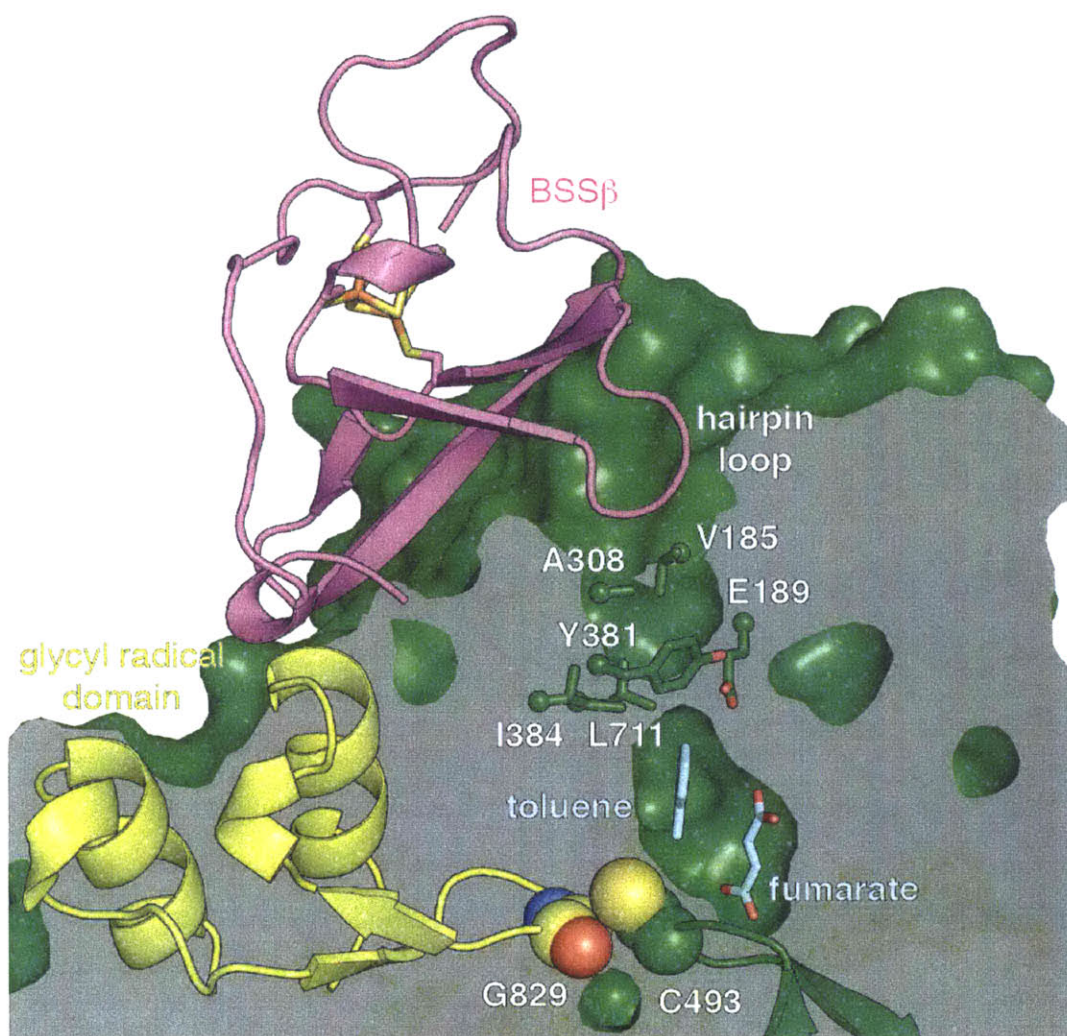


Figure III.3. A putative substrate access channel is gated by protein residues. The van der Waals surface of BSS is shown as a cutaway with substrates (cyan) bound in the central active site cavity above the Cys loop. As was noted in the substrate-free structure (23), BSS β binds both adjacent to the glycy radical domain (yellow) and within a surface cavity that leads to the proposed channel. In particular, a hairpin loop of BSS β inserts into this cavity, filling it completely. The channel also has a bottleneck at the midway point (labeled residues shown as sticks). With BSS β bound, the glycy radical domain appears locked in the active site for catalysis. This structure should represent the catalytically active form, were the glycy radical present.

Fumarate binding stabilizes a partially closed state of BSS α .

We co-crystallized BSS α with 5 mM fumarate to determine if the binding of fumarate alone can convert the open enzyme state, as previously observed in the structure of substrate-free BSS α (23), to the closed state observed in both substrate-free and substrates-bound BSS α β

structures (Figure III.4A). The fumarate-bound BSS $\alpha\gamma$ structure was determined to 2.00-Å resolution, and shows unambiguous density for fumarate in the active site cavity (Figure III.2B). We find that the overall structure of fumarate-bound BSS $\alpha\gamma$ is an intermediate between the closed and open states observed for BSS $\alpha\beta\gamma$ and substrate-free BSS $\alpha\gamma$, respectively (Figure III.4A-C). Instead of generating a fully closed enzyme form, the binding of fumarate to BSS $\alpha\gamma$ stabilizes β 7, β 8, and β 9, which allows a C-terminal helix bundle, which covers the top half of the BSS active site, to adopt a more closed conformation (Figure III.4D). The molecular basis for the remarkable 3-4 Å shift in this region of the protein is revealed by inspection of the active site cavity. In particular, W613 adopts a different rotamer conformation when fumarate is present, and this rotamer shift of W613 allows for β 9 to become ordered (Figure III.5A,B). Fumarate also forms hydrogen bonds with residues on β 7 and β 8, additionally stabilizing both β 8 and β 9. The repositioning and ordering of these β strands allows helices of the C-terminal helical domain to shift into toward the core of the structure (Figure III.5C,D). Although connections between β strands do not typically require stabilization by substrate binding, in BSS β 7, β 8, and β 9 are short and thus only form a small number of inter-strand hydrogen bonds. Moreover, conserved proline residues block some of the hydrogen bonds that would normally be present in a more traditional β sheet or barrel. Given the weakness of the interactions in this region of the protein without substrate, fumarate binding has a large effect in BSS, stabilizing the core barrel, perhaps to prepare the active site to bind toluene. Our previous work showed that BSS β binding stabilizes the barrel in a completely closed conformation that appears more suited to catalysis (23). Thus, both fumarate and BSS β contribute to stabilizing the barrel in the closed conformation, but BSS β causes a more complete closure.

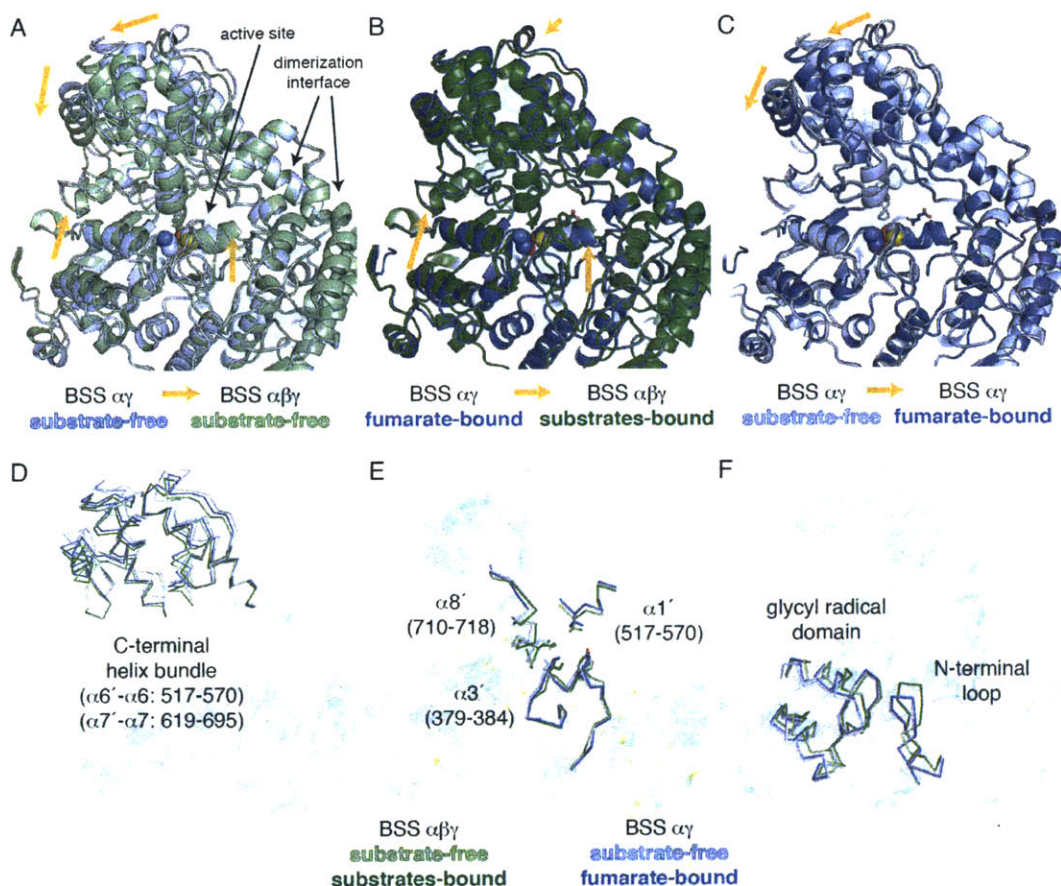


Figure III.4. The structure of BSS α is influenced by the presence of substrates and by the binding of BSS β . (A) Large differences are present between structures of substrate-free BSS $\alpha\beta\gamma$ and substrate-free BSS $\alpha\gamma$ (only one protomer is shown). Arrows indicate a large clamshell motion near the C-terminal helix bundle (see panel D) and glycy radical domain (see panel F) that result in the barrel opening in the absence of BSS β . (B) Between substrates-bound BSS $\alpha\beta\gamma$ and fumarate-bound BSS $\alpha\gamma$, the C-terminal helix bundle is only very slightly shifted. However, the glycy radical domain must still shift inward to form the fully closed conformation that is needed for catalysis. (C) Relative to substrate-free BSS $\alpha\gamma$, the fumarate-bound BSS $\alpha\gamma$ glycy radical domain and associated regions are in the same position, shifted out of the active site. However, the C-terminal helical bundle is shifted toward the center of the protein, leading to a partially closed state. Only BSS α is shown in A-C; there were no changes in the structure of BSS γ . The structures were aligned based on the dimerization interface, which does not change between BSS $\alpha\gamma$ and BSS $\alpha\beta\gamma$ structures. (D) The C-terminal helical bundle is highlighted with all four BSS structures overlaid. Movement of the C-terminal helical bundle away from the glycy radical domain shifts the barrel into the open state. Fumarate-bound BSS $\alpha\gamma$ is an intermediate between the open state found in the substrate-free BSS $\alpha\gamma$ structure and the closed state found in the BSS $\alpha\beta\gamma$ structures. (E) The four structures show varying degrees of openness near three of the helices that comprise the putative substrate access channel (see Figure III.3). (F) The glycy radical domain and an N-terminal loop are highlighted. The loop makes extensive contacts to the glycy radical domain and $\alpha 3'$, and motion of these three elements is correlated in BSS.

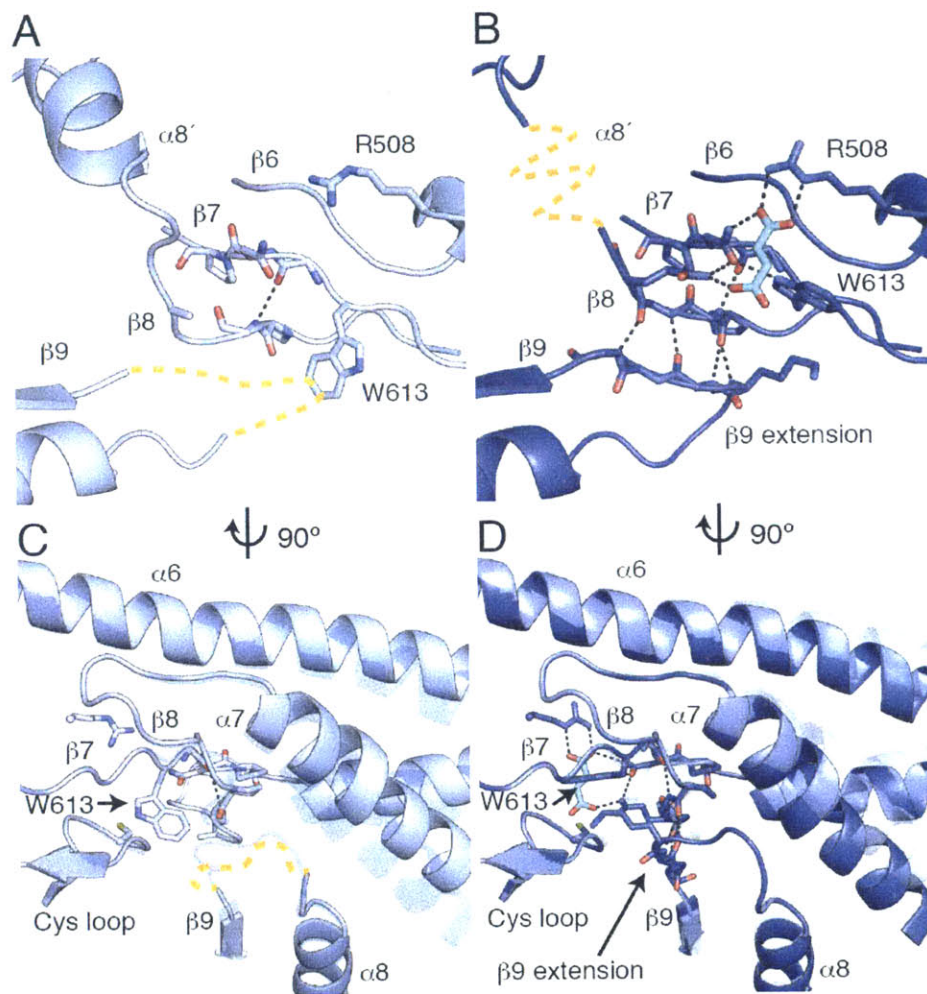


Figure III.5. Fumarate stabilizes the barrel core in BSSay, which in turn allows the C-terminal helix bundle to close in around the core. (A) In the absence of fumarate, contacts between $\beta 7$ and $\beta 8$ are disrupted, and $\beta 9$ is completely disordered (orange dashed line). **(B)** A rotamer flip of W613 enables fumarate binding and enables a chain of conformational changes within the barrel. A network of hydrogen bonds is created between W613, fumarate, and residues on $\beta 7$ and $\beta 8$. Fumarate binding presumably stabilizes $\beta 8$ in a more traditional β sheet conformation, thus allowing an extension of $\beta 9$ to contribute inter-strand hydrogen bonds to $\beta 8$. Conserved proline residues in $\beta 7$ and $\beta 8$ contribute to the especially weak inter-strand contacts in this region. $\alpha 8'$ is disordered in fumarate-bound BSSay (orange dashed line) despite the restoration of the closed barrel (see Figure III.6). **Panel C** and **D** are a side view of panel **A** and **B**. **(C)** The core helices $\alpha 7$, $\alpha 8$, and $\alpha 9$ are shown with a transparent copy of the helices from the other structure for comparison. The rotamer flip of W613 is correlated with movement and disorder in $\beta 9$ and the $\beta 9$ extension (orange dashed line), as the indole side chain would clash with backbone residues in this conformation. The movement in $\alpha 7$ is likewise correlated with the changes in $\beta 9$ extension, as these residues come in close contact with each other. **(D)** Fumarate binding stabilizes the positions of W613 and $\beta 9$ in the closed state.

Despite this large, fumarate-induced shift in the barrel, the active site in fumarate-bound BSS $\alpha\gamma$ remains open to solvent due in part to the unfolding of helix $\alpha 8'$ (residues 710-718) (Figure III.6A) as well as a shift in the position of helices $\alpha 3'$ (residues 379-384) and $\alpha 1'$ (residues 176-184) (Figure III.4E) away from the glycyl radical domain. Helix $\alpha 8'$ residues are also poorly ordered in the substrate-free BSS $\alpha\gamma$, although it was possible to model the residues based on the structure of BSS $\alpha\beta\gamma$ in this case (Figure III.6B). Interestingly, toluene soaks of fumarate-bound BSS $\alpha\gamma$ crystals did not lead to a toluene bound structure, suggesting that this open state is not the form of the enzyme to which this second substrate can stably bind.

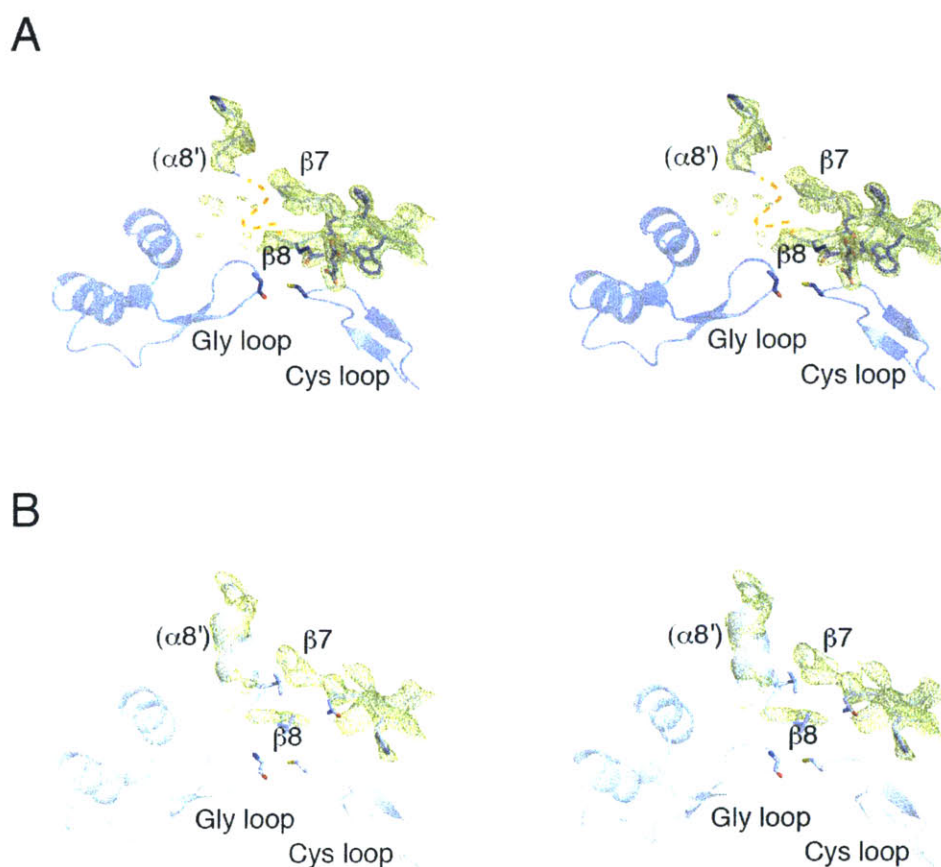


Figure III.6. Helix $\alpha 8'$ is disordered in BSS $\alpha\gamma$ structures. (A) Stereoimage showing $2mF_o - DF_c$ electron density contoured at 1σ around the non-core helix $\alpha 8'$ and the core strands $\beta 7$ and $\beta 8$. There is density present suggesting that helix $\alpha 8'$ is present but structurally heterogeneous. The missing residues have been marked (orange dashed line) **(B)** In substrate-free BSS $\alpha\gamma$, $\alpha 8'$ is better ordered despite deterioration in the density for $\beta 7$ and $\beta 8$.

Unlike the C-terminal helix bundle, binding of fumarate does not change the position of glycy radical domain and associated N-terminal loop relative to substrate-free BSS $\alpha\gamma$. These regions remain shifted out of the active site by ~ 2 Å (Figure III.4F) relative to their position in BSS $\alpha\beta\gamma$. There are no contacts between fumarate and the glycy radical domain, and it appears that the partial closure of the barrel between $\beta 8$ and $\beta 9$ is independent of the outward shift of the glycy radical domain.

BSS lines up fumarate and toluene for radical chemistry

We previously noted features of the BSS active site that could be involved in binding substrates (23). In particular, we predicted that a chloride ion that was found to interact with backbone amides of the Cys loop might be occupying the binding site of one of the fumarate carboxylates (Figure III.7A), with the other carboxylate of fumarate interacting with a conserved arginine residue from $\beta 6$ (R508). Structures of both BSS $\alpha\gamma$ and BSS $\alpha\beta\gamma$ show identical fumarate coordination and confirm that these features do indeed contribute to the binding of fumarate. The proximal carboxylate of fumarate does hydrogen bond to the backbone amides of Cys loop (NH of residues L492, M494, and S495), replacing the chloride ion observed in the substrate-free structure (Figure III.7B), and R508 does anchor the distal carboxylate of fumarate in a bidentate salt bridge (Figure III.8A). Fumarate is held nearly planar, with contacts to W613 in back and L492 in front ensuring the *re* face of the C2 carbon of fumarate (see Scheme III.1 for numbering) is facing out toward toluene and C493 (which becomes the putative, transient thiyl radical). A hydrogen-bonding network that is formed by residues N615 and Q707 (on $\beta 7$ and $\beta 8$, respectively) and fumarate anchors both carboxylates of the substrate, securing fumarate into position at the back of the active site. W613 joins this hydrogen-bonding network with an indole nitrogen contact to N615 but also forms an edge-to-face interaction with the double bond of fumarate. The enzyme actively discriminates against dicarboxylates other than fumarate: modeling of succinate in the active site reveals steric clashes with L492 and W613 (Figure III.9). Accordingly, even after soaking crystals of BSS $\alpha\gamma$ in 100 mM succinate, no density was observed in the active site.

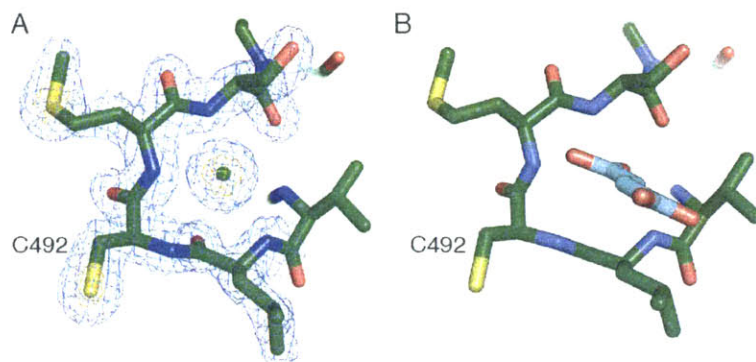


Figure III.7. Active site environment in BSS. (A) Reproduced from (23): "Cys loop with a chloride binding site in the BSS $\alpha\beta\gamma$ structure. $2mF_o-DF_c$ density contoured at 1σ (blue mesh) and 5σ (orange mesh) is consistent with a chloride ion (green sphere) which coordinates the four amide nitrogens of residues 492-495 with Cl-N distances between 3.12-3.50 Å." **(B)** BSS $\alpha\gamma$ Cys loop with fumarate present. The proximal carboxylate group binds in the space previously occupied by the chloride anion.

Toluene contains no polar functional groups to facilitate its positioning during catalysis. The binding site is thus created primarily by aliphatic and aromatic residues. Residues from two regions, $\alpha 3'$ (I384, F385, L391) and the loop between $\beta 8$ and $\alpha 8'$ (V709, L711), create what we have called a "hydrophobic wall" that forms one side of the active site (Figure III.8B), opposite fumarate. A ring of hydrophobic (Y197, Y381, L492, I617) and polar (E189, Q707) residues forms the sides of the toluene binding pocket, matching the contours of toluene perfectly (Figure III.8C). The final side of the binding pocket is created by fumarate itself, with the methyl of toluene stacked against C2 of fumarate and the aromatic ring of toluene stacked against the distal carboxylate of fumarate. This orientation of toluene in the pocket is consistent with the production of (*R*)-benzylsuccinate, as previously observed (25, 26), since it places C1 of toluene directly adjacent to C2 of fumarate (Figure III.8D). The observed distances between C α of G829 (the putative glycy radical), S γ of C493 (the putative transient thiyl radical) and the C1 of toluene are 3.4-3.9 Å, consistent with observed donor-acceptor distances in a number of other radical systems (27). The two substrates are thus pre-arranged to permit C-C bond formation with little to no movement.

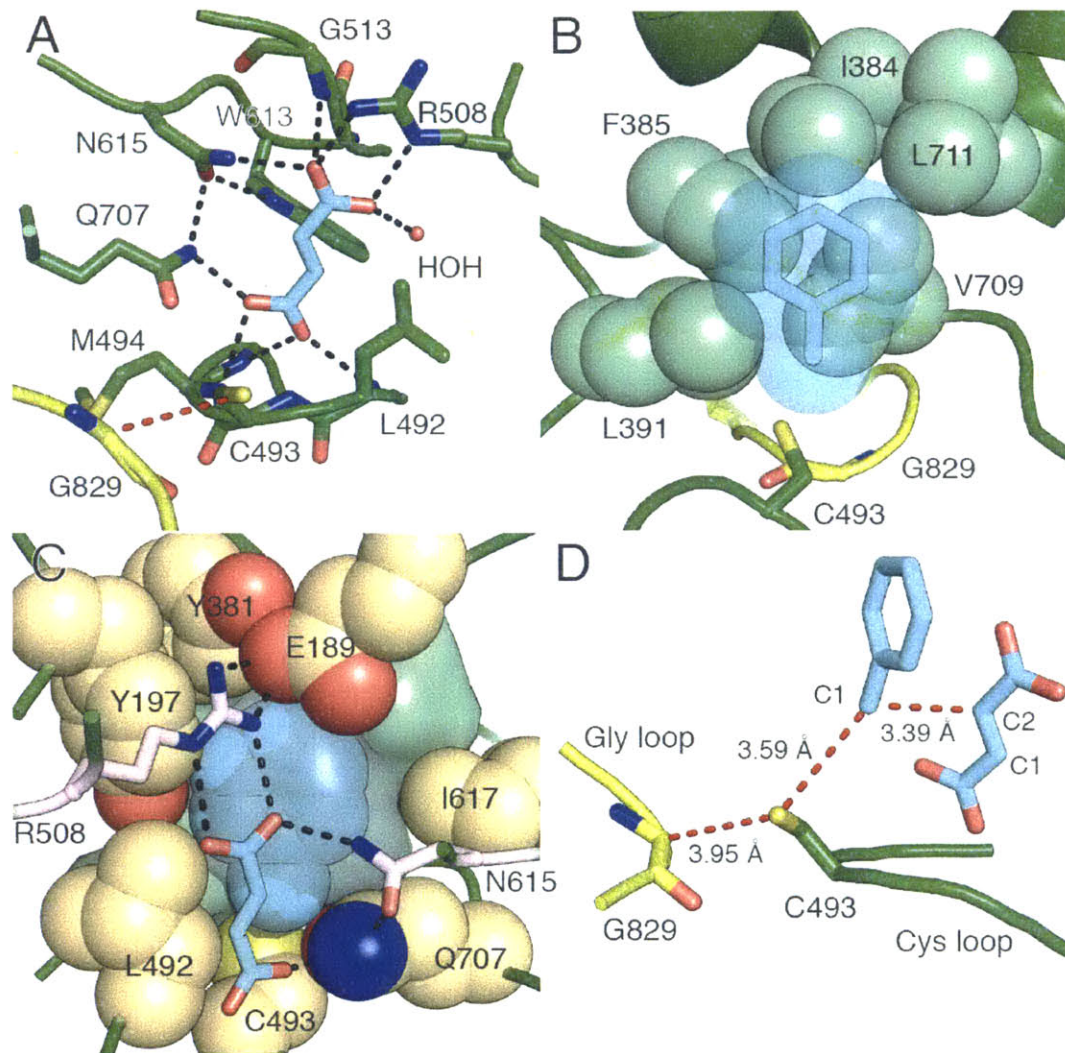


Figure III.8. Substrates bind BSS adjacent to the active site Cys loop. (A) BSSay (green) with fumarate (cyan) bound above the Cys loop. The glycy radical domain (yellow) is shown with the radical transfer path (red dashes) between the G829 and C493. Hydrogen bonds (black dashes) are shown for polar distances less than 3.2 Å. A single water molecule (HOH) participates in fumarate binding. **(B)** Van der Waals sphere representation of the hydrophobic wall (light green spheres) involved in binding toluene (cyan transparent spheres and sticks). The orientation is 180° along y from that in panel A. **(C)** The toluene binding site is shown with residues contributing to the sides of the binding pocket (tan spheres). The hydrophobic wall is present behind toluene (light green surface). Fumarate (cyan sticks) is shown in the foreground with residues that provide hydrogen bonds and contribute to the toluene pocket (pink sticks). The cavity contours around toluene are shown (gray surface) to illustrate the tight packing. **(D)** The toluene/fumarate co-complex positions substrates such that the expected radical transfer distances (red dashes) are minimized. The orientation is 90° along y from that in panel A.

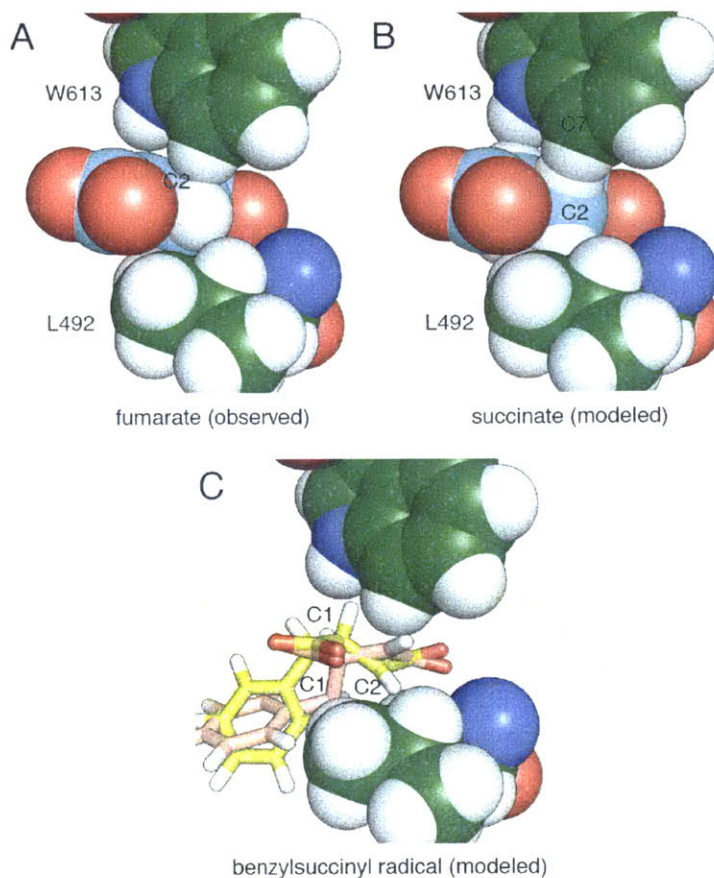


Figure III.9. The BSS active site discriminates against non-planar dicarboxylates. (A) W613 and L492 in the BSS α y structure pack against C2 of fumarate with an edge-to-face interaction and van der Waals contacts, respectively. **(B)** Modeled succinate clashes with both residues through hydrogen atoms. **(C)** Two different conformations of the *R*-benzylsuccinyl radical modeled into the fumarate-bound active site clash with either W613 (yellow) or L492 (pink).

Two active site regions contribute to specificity in X-succinate synthases.

Residues within the Cys loop that contribute to fumarate binding are conserved within putative XSSs but not in other GRE families (Figure III.10). 4-Isopropylbenzyl-SS (IBSS), 4-hydroxybenzyl-SS (HBSS), and naphthyl-2-methyl-SS (NMSS) are essentially identical across the aryl-SSs, but the more divergent 1-methylalkyl-SS has some substitutions in the Cys loop that may have catalytic relevance (discussed below). However, there is much less sequence conservation within residues that are predicted to determine hydrocarbon specificity (Figure III.11).

```

BSS      NGATDEEAHNWVNVLCMSSPGIHGRR-----KTQKTRSEG 511
IBSS     NGATPEEAHDWANVLCMSSPGLVGRR-----KTQKTRSEG 506
HBSS     ---PEEWARDWALVLCMSSPGITGRR-----GTQKTRSEG 468
NMSS     ---PPDEAAHWALVLCMAPGVGKRR-----GLQKTRTEG 504
MASS     HGHPLEEARWVHQACMSSPCP*TKH-----GFQPF*RMAS 515
HPD      EGMDLHDARAWCLGGCLESAPGCFLPLEYNGKVTMIPGGASPTCGTG 534
PFL2     AEVDDRDALNYTTDGCVEIAPFGNS-----F----TSSD 440
GDH      RGLTLEDARDYGIIGCVEPQKPGKT-----E----GWHD 447
CutC     KGV*DIEDARDYCLMGCVEPQKSGRL-----Y----QWTS 505
PFL      ---PDFNDDYAIACCVSPMIVGKQ-----MQFFGARAN
          :      * :

```

Figure III.10. Alignment of Cys loop and adjacent residues in GREs. XSS enzymes (top five sequences) are distinguished by the presence of a CM[AS]P motif followed by an absolutely conserved arginine (R508) that contacts fumarate directly (see Figure III.8A). The Cys loop in BSS is underlined and conserved residues in the active site are in bold. The alignment was performed in Clustal Omega (28). BSS: benzylsuccinate synthase, IBSS: isopropylbenzylsuccinate synthase, HBSS: hydroxybenzylsuccinate synthase, NMSS: naphthyl-2-methyl-succinate synthase, MASS: (1-methylalkyl)succinate synthase, PFL2: pyruvate formate-lyase-like protein from *Archeoglobus fulgidus*, GDH: glycerol dehydratase from *Clostridium butyricum*, CutC: choline trimethylamine-lyase from *Desulfovibrio alaskensis*, HPD: 4-hydroxyphenylacetate decarboxylase from *Clostridium difficile*, PFL: pyruvate formate-lyase from *Escherichia coli*.

To assess how the residues of the hydrophobic wall could contribute to substrate specificity we constructed homology models of four characterized XSS family members based on the structure of BSS $\alpha\beta\gamma$ (Figure III.12A) and manually docked their known substrates into the models. IBSS is closest in sequence conservation to BSS (72% identity) and has only two substitutions (L711V, I384A) out of 13 active site residues (Figure III.12B). Both residues are found at the far end of the active site, away from the site of hydrogen atom abstraction and directly adjacent to the isopropyl group of the modeled substrate, *p*-cymene (4-isopropyltoluene). The result of these two substitutions is an enlarged binding site tailored perfectly for the desired substrate. Similarly, HBSS (54% identity) contains four substitutions, which cluster at the end of the active site near the phenolic oxygen of the modeled substrate, *p*-cresol (Figure III.12C). The net effect of these substitutions appears to be a slightly larger binding site for *p*-cresol in HBSS than for the smaller toluene in BSS. A unique E189Q substitution in the HBSS subfamily would appear to place a glutamine adjacent to the phenolic oxygen of the substrate. These differences between BSS and HBSS are not dramatic, and it is therefore not surprising that BSS also accepts *p*-cresol as a substrate, although with reduced affinity relative to toluene (29).

alkyl specificity (cyan), and other residues that contribute to hydrocarbon specificity (green). Residues conserved between BSS and related enzymes are also highlighted. Regions of the alignment particularly important for substrate binding (bold) and the active site glycine and cysteine residues (red) are indicated. The alignment was performed in Clustal Omega (28). Abbreviations are as in Figure III.10.

NMSS (49% identity) has six substitutions (E189D, I384G, F385N, L711T, I617P), all of which decrease the size of the amino acid side chains in the active site, thus allowing the second aryl ring of the substrate, 2-methylnaphthalene, to fit in the cavity (Figure III.12D). In contrast to the top face of the active site, which varies substantially between different enzymes, aryl-SSs have a core set of conserved residues at the bottom of the active site, which position the methyl group of toluene between S γ of C493 and C2 of fumarate in BSS. This pattern of conservation suggests that these enzymes maintain the same substrate geometry around the methylbenzene moiety of the substrate in order to perform the required chemistry. Thus, hydrocarbon specificity appears governed by substitution of residues at the top of the active site, whereas residues at the bottom of the active site involved in fumarate positioning are conserved.

A distinct subset of XSS enzymes has been proposed to generate various 1-methylalkylsuccinate products instead of aryl-succinates. MASS (35% identity to BSS) is likely too diverged to model with confidence in precise side chain placement, but key differences from the other aryl-SS enzymes are still readily apparent. Only two residues that make up the toluene-binding pocket in BSS are conserved in MASS (V709 and Y197), and three residues crucial to the binding of toluene in BSS (L711, I384, E189) are replaced by glycines, suggesting the active site may be completely remodeled to bind to linear alkyl chains such as the modeled *n*-hexane (Figure III.12E). Although MASS retains two key residues for binding fumarate in BSS, R508 and W613, it is notable that there are substitutions in other residues (G513A, Q707E, N615T) that contribute to binding of fumarate (Figure III.8A). The disruption of these important contacts, and especially substitution at glycine, may indicate that large structural changes in the architecture of the MASS active site reposition fumarate within the active site cavity to allow for its interaction with a non-aryl substrate.

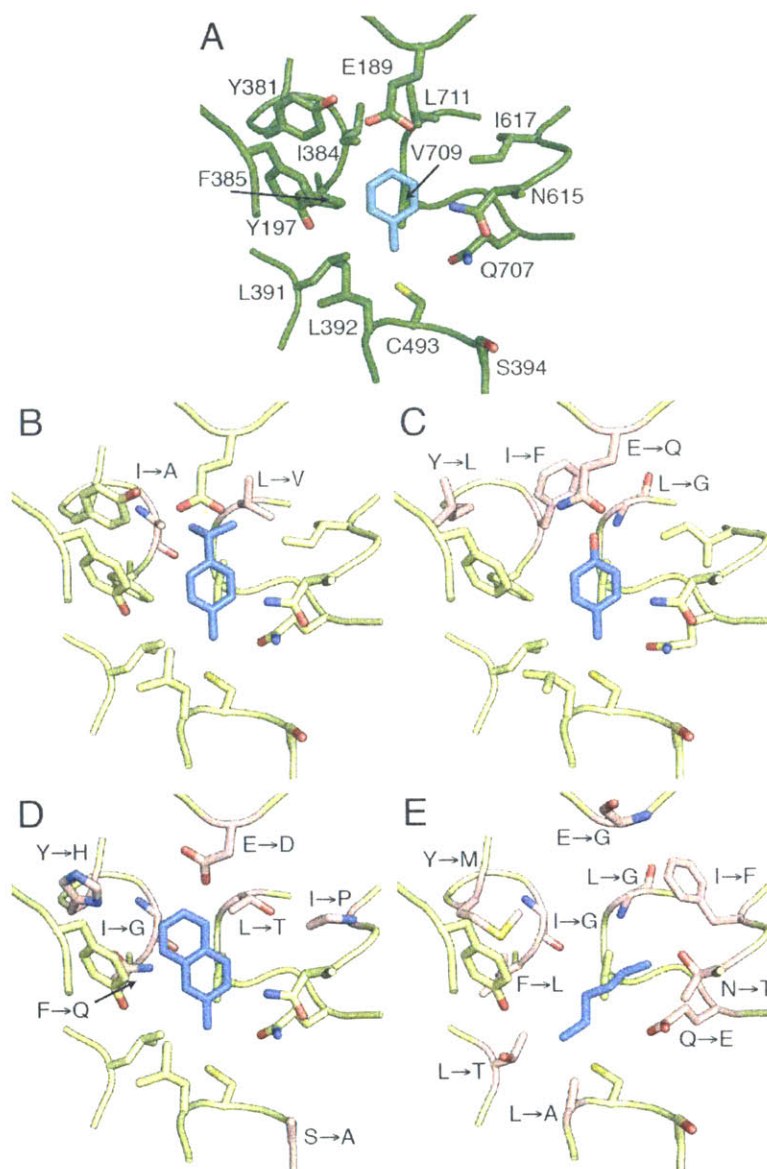


Figure III.12. Homology models of XSS family members displaying the hydrocarbon binding site. (A) Crystal structure of BSS $\alpha\beta\gamma$ (green) with toluene bound (cyan). **(B)** Model of IBSS with *p*-cymene bound. **(C)** Model of HBSS with *p*-cresol bound. **(D)** Model of NMSS with 2-methylnaphthalene bound. **(E)** Model of MASS with *n*-hexane bound. Models are in yellow with manually modeled substrates in blue. Substitutions are indicated by one-letter code and are colored pink.

III.D Discussion

Although GREs perform a diverse set of chemical transformations, they utilize a common architecture for radical storage and to enable handling of radical intermediates. These enzymes are faced with the challenge of alternatively needing to adopt open and closed barrel states. The

open state is essential to expose the Gly loop in order to install the radical within unactivated enzymes. The enzyme must then return the barrel to the closed state to protect the newly formed glycy radical and the subsequent radical intermediates thought to be generated during catalysis. In addition, XSS enzymes have the special task of securing hydrocarbons like toluene or *n*-hexane that have no functional handles. These molecules must be held in the active site with precision so that hydrogen atom transfer can take place, while also allowing for an access channel to open for substrate-binding and product release. Thus, closed and open states of the putative substrate-access channel are also required once the enzyme has been activated.

XSS enzymes are unusual among GREs in their ability to bind two small subunits, BSS β and BSS γ , both of which appear to directly impact the stability of the protein (20, 23). BSS γ is most important for the solubility of BSS α , whereas BSS β is proposed to regulate the conformational dynamics of BSS α . In particular, we previously showed that the absence of BSS β leads to an unfolding of the active site barrel such that the glycy radical domain can exit and access to the active site for substrates is not hindered, whereas in the presence of BSS β the active site is sealed. To determine if substrate binding to the BSS β -free complex leads to a conformational change that seals the active site, we determined a fumarate-bound structure of BSS $\alpha\gamma$. Although we find that fumarate binding does lead to a more compact structure, the resulting partially closed barrel is not ready for catalysis since regions of the protein important for toluene binding, including the hydrophobic wall, are still disordered in this structure. Further, we were unable to capture a structure with toluene bound to BSS $\alpha\gamma$, suggesting that the closing of the channel and ordering of the hydrophobic toluene binding pocket, as observed in the BSS $\alpha\beta\gamma$ structure, may be necessary to retain toluene in the active site. Additionally, the binding of fumarate alone to the open BSS $\alpha\gamma$ complex does not change the position of the glycy radical domain or shorten the distance between the active site Gly and Cys loops, suggesting that extra factors besides fumarate-binding are necessary to generate a form of the enzyme capable of radical transfer. One of these factors may be activation itself, which is expected to change the backbone conformation of the Gly loop and thus may allow for deeper packing of the glycy radical domain within the active site. Alternatively, the binding of BSS β may be the trigger that is needed for both the correct positioning of the glycy radical domain and for the formation of the toluene binding pocket, readying the enzyme for catalysis.

Prior to this work, it was not clear if the BSS $\alpha\beta\gamma$ complex would be able to bind substrates as the channel appears to be blocked in the presence BSS β . Here we find that substrates do bind to BSS $\alpha\beta\gamma$ and appear to be positioned perfectly for catalysis. In retrospect, it makes sense that toluene could only be stably bound in an active site that is largely sealed. With no functional groups to hold on to, the enzyme must create a binding pocket around the toluene to position it for catalysis, which is what we observe. Although it seems that some rearrangement or partial dissociation of BSS β must occur to allow for substrate access, these structures suggest that BSS β will be bound during enzyme turnover.

The order of substrate binding in BSS has not yet been determined; however, there are numerous reasons to believe that fumarate will bind before toluene. First, fumarate lies deep within the active site of BSS, whereas toluene binds above it, adjacent to residues thought to form the channel (Figure III.2). Second, our structures show that in BSS $\alpha\gamma$ fumarate can bind in the absence of toluene and that fumarate binding does not completely order the active site in this complex. Third, fumarate forms one side of the toluene binding pocket, indicating that tight binding of toluene requires the presence of fumarate (Figure III.8C). Finally, deuterium-labeling studies have shown that hydrogen atom abstraction first occurs from C1 of toluene (25, 26), so obligate binding in this order would prevent off-pathway production of the free benzyl radical, which could damage the enzyme.

Taken together, our data allow us to propose a structure-based mechanism (Figure III.13) that builds on proposals from previous biochemical and computational experiments (30, 31). We show fumarate binding first (Figure III.13, i) and toluene binding second (ii), for the reasons discussed above. Once toluene enters the active site, L492, Q707, and residues of the hydrophobic wall position the C1 of toluene perfectly between S γ of C493 and C2 of fumarate. Hydrogen atom abstraction from C1 of toluene can therefore result in immediate formation of the C–C bond between toluene and fumarate as the singly occupied p-orbital of the benzyl radical is already adjacent to the double bond of fumarate (iii-iv). The resulting benzylsuccinyl-C3-radical intermediate would need to change orientation in the active site to regenerate the thiyl radical at C493 (v-vi). This change in orientation may be facilitated by the adjacent L492 and W613, both of which pack tightly against fumarate. The loss of planarity at C2 after C–C bond formation and movement of the C2 hydrogen toward W613 may force the proximal carboxylate to shift, changing the orientation of the benzylsuccinyl intermediate and thus facilitating hydrogen atom

abstraction from C493 by the C3 radical. The exit of benzylsuccinate (vi-vii) may be facilitated by L492 and W613, which are arranged in the fumarate-bound structure such that they should form unfavorably close interactions with the sp^3 hybridized C2 and C3 of the product, benzylsuccinate.

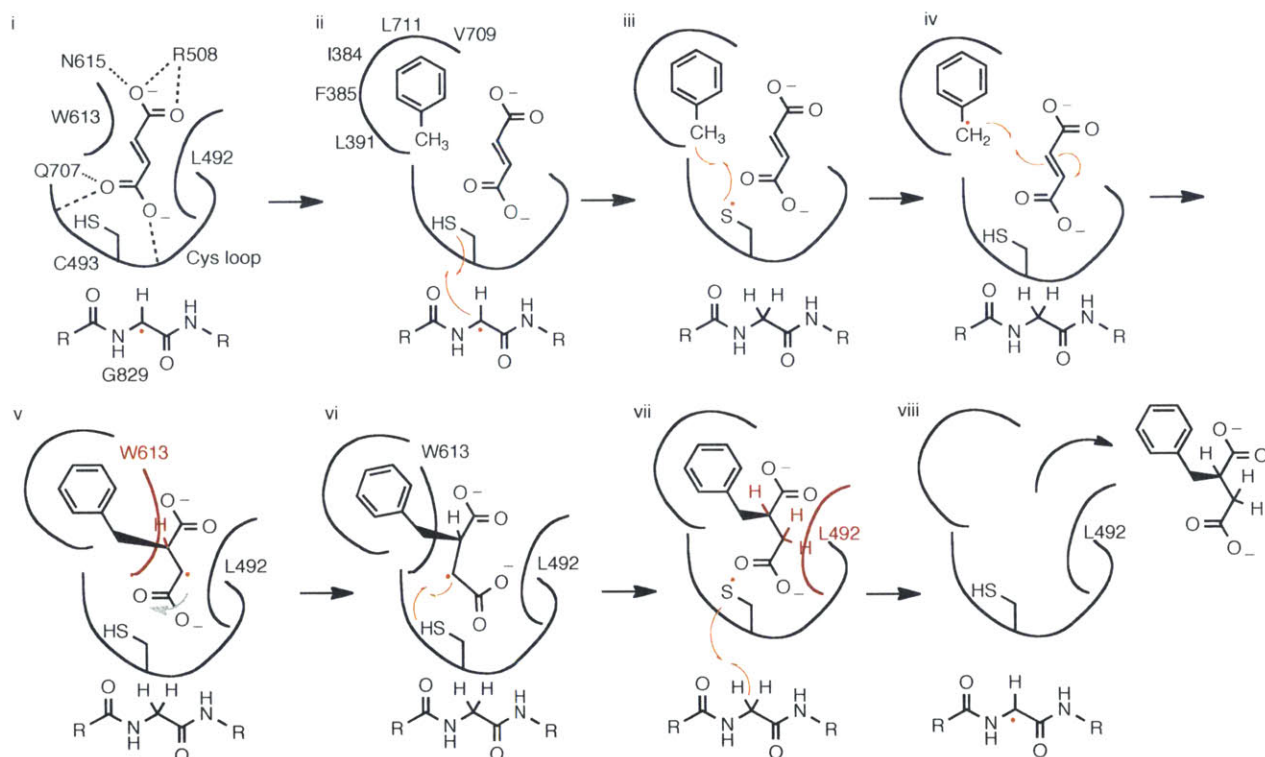


Figure III.13. Structure-based mechanistic proposal for BSS. (i) Conserved residues facilitate fumarate binding above the Cys loop. (ii) Toluene binds following fumarate; binding and specificity are effected by the hydrophobic wall and surrounding residues. Binding of both substrates facilitates hydrogen atom transfer from C493 to the G829 glyceryl radical cofactor. (iii) The transient thiyl radical abstracts a hydrogen atom from toluene to generate a benzyl radical. (iv) C–C bond formation is facilitated by pre-organizing substrates directly adjacent to each other. (v) The putative benzylsuccinyl radical intermediate clashes with W613 causing a rearrangement that places C3 close to C493. (vi) Hydrogen atom re-abstraction from C493 forms benzylsuccinate. (vii) Hydrogen atom re-abstraction from G829 regenerates the radical storage cofactor. Clashes with L492 and W613 help eject benzylsuccinate. (viii) Once the radical has returned to the Gly loop, the product is free to exit the active site.

Our structures of BSS with substrates bound provide a fresh prospective on sequence motifs of putative XSS enzymes in genomic databases. Specifically, a Cys loop sequence motif of LCM[AS]P appears to be diagnostic of aryl-SSs, whereas an [AS]CMSP motif would seem to specify an alkyl-SSs (Figure III.10). In both of these classes of XSS enzymes, the small side

chains of alanine or serine two residues after the active site cysteine create space for fumarate to bind to the backbone amides of the Cys loop. Instead of alanine or serine, a glutamate is commonly found in this position in GREs that perform dehydration reactions and is thought to participate in catalysis as a general base (32). The conserved methionine and proline residues are likely important for the overall structure of the Cys loop as similar residues are observed in non-XSS GREs (Figure III.10).

We suspect that the difference in the amino acid identity one residue upstream of the cysteine in the Cys loop—leucine for aryl-SSs and alanine for alkyl-SSs—has to do with positioning of the substrate. In aryl-SS enzymes, the larger leucine positions the aryl substrate further from cysteine, allowing for C1 hydrogen atom abstraction. The initial hydrogen atom abstraction from alkanes by MASS is predicted to be substantially more difficult than that proposed for aromatic hydrocarbons (33) due to the lack of resonance delocalization in the product-based radical. The growth rate of alkane-degrading species is notably much slower than related aromatic-hydrocarbon-degrading species (34), perhaps in part due to the difficulty of the initial activation reaction, in which a saturated hydrocarbon C–H bond is broken rather than the moderately activated aryl C–H bond. The elevated barrier for this hydrogen atom abstraction may be somewhat mitigated by performing radical abstraction at the subterminal methylene carbon (C2) rather than the terminal methyl (C1). For the reaction to occur with this regiochemistry, the methyl group of the alkyl chain must be positioned differently such that C2 is adjacent to the active site cysteine. The substitution of alanine for leucine could make room in the active site for the terminal methyl group in alkyl-SS enzymes. There have also been proposals of terminal hydrogen atom abstraction in short chain alkanes (35) and even the proposal that methane could be activated by an XSS (36). However, the ability of an alkyl-SS to perform such a dramatically uphill C–H bond cleavage (37) has not been substantiated in vitro to date, and no candidate XSS has been proposed for this reaction.

Within BSS, residues in the Cys loop are important for fumarate binding, the stereochemistry of the reaction, and the reaction mechanism, which is likely to be conserved in related aryl-SSs. However, residues at the upper end of the active site appear to vary considerably and thus govern substrate specificity. BSS, HBSS and IBSS have toluene or toluene-like substrates, and are similar enough in sequence that homology modeling allows us to propose residue substitutions that could account for their different substrate specificities. The

much lower sequence identity of NMSS with BSS makes rationalizing its substrate preference less certain, but a conservative interpretation is that residue substitutions at the top of the active site and along the hydrophobic wall serve to enlarge the active site to accommodate the larger substrate of NMSS. The XSS family member MASS shows the greatest divergence in sequence, which is consistent with its role as an alkyl-SS as these substrates require a change not only in the hydrocarbon binding pocket, but also in the region of the active site that governs radical transfer specificity. However, the very low sequence identity to BSS within the active site strands and loops of MASS makes drawing specific inferences about the chemistry occurring within alkyl-SS enzymes untenable. Although these homology models are useful for proposing general trends in active site architecture, detailed biochemical and structural investigation is needed for more members of the XSS family in order to fully understand how this class of GREs recognizes substrates and performs radical chemistry.

The structures of BSS with bound substrates presented here greatly advance our knowledge of a key reaction in the anaerobic degradation of hydrocarbons. We have identified residues that may determine specificity in XSS enzymes and thus provide a genetic handle for determining the hydrocarbon degrading potential of microbial communities. Although hydrocarbon degradation is a beneficial process in natural environments, in industrial settings oil fouling by microbes can be detrimental to oil quality as well as to physical infrastructure. Inhibitors of anaerobic hydrocarbon degradation inspired by these structures could therefore be useful additives. These structures also provide a starting place for future efforts to characterize the BSS reaction computationally, especially to investigate the series of rearrangements occurring after C–C bond formation that return the radical to its storage site within the protein backbone.

III.E Materials and Methods

BSS $\alpha\beta\gamma$ and BSS $\alpha\gamma$ were purified (20) and crystallized (23) as previously reported. Crystallization was performed by sitting drop vapor diffusion at room temperature. All crystallization reagents were from Hampton Research. Initial crystals were discovered using screens dispensed by a TPP Labtech Mosquito liquid-handling robot. BSS $\alpha\beta\gamma$ at 8 mg/mL in buffer containing 50 mM Tris pH 7.6, 15% (v/v) glycerol, and 200 mM NaCl was added at a 2:1 ratio to well solution containing 25% (w/v) PEG 3350, 100 mM Tris pH 8.5, 60 mM KCl,

and 5 mM fumarate. 1-2 μL toluene was added to the bottom of the well and allowed to diffuse slowly into the protein drop. Diffraction-quality crystals grew over the course of three weeks. Crystals were soaked in a cryoprotection solution containing 50 mM Tris pH 8.5, 25% (w/v) PEG 3350, 10% (v/v) glycerol, 50 mM fumarate for 30 s before cryocooling. Crystals indexed in space group $P2_12_12$ ($a = 141.5$, $b = 115.4$, $c = 121.7$ Å) with two molecules per asymmetric unit. This crystal form is different from that observed before for substrates-free BSS $\alpha\beta\gamma$ (space group I222; $a = 113.4$, $b = 120.4$, $c = 136.0$ Å), but the packing interactions are virtually identical, with small changes accounting for the lower crystallographic symmetry. BSS $\alpha\gamma$ at 15 mg/mL in buffer containing 20 mM HEPES pH 7.6, 100 mM NaCl, and 5 mM fumarate was mixed in a 1:1 ratio with precipitant solution containing 20% (w/v) PEG 400, 50 mM bis-Tris pH 6.5, 25 mM Tris pH 8.0 and incubated over a well of the same solution at room temperature. Crystals formed in several days and were pale yellow. Crystals were soaked in a solution containing 30% (w/v) PEG 400, 50 mM bis-Tris 6.5, 25 mM Tris 8.0, and 50 mM sodium fumarate for 5-10 min before cryocooling. Crystals indexed in space group $P4_32_12$ ($a = b = 154.9$, $c = 82.1$ Å) with one molecule per asymmetric unit. All crystallization steps were performed in a room-temperature, N_2 -filled, MBraun anaerobic chamber with $\text{O}_2 < 0.1$ ppm. Crystals were cryocooled in liquid nitrogen in a Coy anaerobic chamber filled with 95% Ar, 5% H_2 .

All data were collected at the National Synchrotron Light Source beamline X26C on a Quantum-4 ASDC CCD detector at a wavelength of 1.000 Å (Table 1). Data were indexed and scaled in HKL2000 (38). Resolution cutoffs were chosen at a $\text{CC}_{1/2}$ value of ~ 0.7 , consistent with the deterioration of the signal to noise and R_{cryst} beyond this point. The structure of fumarate-bound BSS $\alpha\gamma$ was solved initially at 2.3-Å resolution by molecular replacement using the Phenix implementation of Phaser (39) with the previously solved structure of BSS $\alpha\gamma$ as a model (23). Manual rebuilding of the model was performed in Coot (40) followed by refinement with phenix.refine (41) at 2.0-Å resolution. Translation/libration/screw refinement of B factors improved R_{free} substantially. The final model contains residues 6-708, and 716-865 (out of 865 native residues) in BSS α and 9-47 (out of 60) in BSS γ . Residues 709-715, which comprise part of $\beta 8$ and $\alpha 8'$, are disordered. The 14 residue C-terminal linker and His $_6$ -tag are not observed in BSS α . The model contains 97.1% of residues in the most favorable region of the Ramachandran plot, 2.7% in additionally allowed regions, and 0.2% (4 residues) in outlier regions. As in the previously reported structure (23), the [4Fe-4S] cluster and three of the four cysteine residues

coordinating it from BSS γ are not observed in the electron density. Parameter files for fumarate (ligand ID FUM) were obtained from the CCP4 database.

Table 1. Data processing statistics

	fumarate- and toluene- bound BSS $\alpha\beta\gamma$	fumarate-bound BSS $\alpha\gamma$
Space group	P2 ₁ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions (Å)	141.5, 115.4, 121.7	154.9, 154.9, 82.1
Beamline	NSLS ¹ X26C	NSLS ¹ X26C
Wavelength (Å)	1.0000	1.0000
Resolution (Å)	40.0-3.30 (3.36-3.30)	35.0-2.00 (2.03-2.00)
Completeness %	98.9 (95.4)	97.0 (80.4)
<I/ σ >	9.6 (1.3)	11.8 (1.4)
R _{sym}	0.101 (>1)	0.085 (0.900)
Unique reflections	30311	65770
Redundancy	3.8 (3.5)	5.6 (2.8)
CC _{1/2}	(0.669)	(0.700)

¹National Synchrotron Light Source, Brookhaven National Laboratory

Table 2. Model statistics.

	fumarate- and toluene- bound BSS $\alpha\beta\gamma$	fumarate-bound BSS $\alpha\gamma$
Model refinement		
Resolution (Å)	38.3-3.30	34.6-2.00
R _{work} , R _{free}	0.212, 0.237 ¹	0.195, 0.223 ²
Protein atoms	15350	7063
Toluene atoms	14	NA
Fumarate atoms	16	8
Other ligand atoms	16 ³	71 ⁴
Water molecules	0	415
Bond length deviation (Å)	0.003	0.003
Bond angle deviation (°)	0.913	0.705
Rotamer outliers (%)	0.37	0.13
Average B factors (Å ²)		
BSS α	111.1	53.7
BSS β	128.6	NA
BSS γ	130.0	53.9
Water atoms	NA	42.8
Toluene	95.2	NA
Fumarate	94.2	41.6
Other ligands	130.3	61.0

¹6% of reflections (1809) were set aside for cross validation.

²5% of reflections (3271) were set aside for cross validation.

³two [4Fe-4S] clusters are present within BSS β

⁴six polyethylene glycol molecules and a molecule of Tris buffer are present at the protein surface

The structure of toluene- and fumarate-bound BSS $\alpha\beta\gamma$ was solved by molecular replacement at 3.30-Å resolution in the Phenix implementation of Phaser using the published chloride-bound BSS $\alpha\beta\gamma$ structure (23) as an initial model. Due to the low resolution of this dataset, refinement was tightly restrained with the high-resolution model as a reference structure in phenix.refine. The model contains residues 9-865 (out of 865 native residues) in BSS α , 13-81 (out of 81) in BSS β , and 11-47 (out of 60) BSS γ . The model contains 97.4% of residues in the most favorable region of the Ramachandran plot, 2.2% in additionally allowed regions, and 0.4% (8 residues) in outlier regions. The 14 residue C-terminal linker and His₆-tag are not observed in BSS α . Composite omit maps calculated by phenix.autobuild were used to verify positions and

identity of ligands and active site residues in both structures. Structure figures were generated in PyMOL v1.4.1 (Schrödinger, LLC). Parameter files for toluene (MBN) were obtained from the CCP4 database.

Homology models of the catalytic subunit of IBSS (GenBank ID AIS23708.1), HBSS, (CCK78655.1), NMSS (ADB04297.1), and MASS (CAO03074.1) were constructed with Modeller v9.1 (42) using the structure of BSS $\alpha\beta\gamma$ as a starting model. The models have Discrete Optimized Protein Energy scores ranging from -96600 to -10700. Ligand models for docking were prepared in phenix.eLBOW. Docking was performed manually based on the position of toluene in the active site of BSS $\alpha\beta\gamma$.

III.F Acknowledgements

E. N. G. Marsh generously provided plasmids containing the *tutFD* operon and *tutG*. This work was supported in part by the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. 0645960 (M.A.F.) and NSF Grant MCB-0543833 (C.L.D.). C.L.D. is a Howard Hughes Medical Institute (HHMI) Investigator. Data were collected at Beamline X26A, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. X26A is supported by Department of Energy (DOE) Geosciences (DE-FG02-92ER14244 to the University of Chicago Center for Advanced Radiation Sources). Use of the NSLS was supported by DOE Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886.

III.G References

1. Leahy JG & Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol Rev* 54(3):305-315.
2. Board MBaOS (2003) *Oil in the Sea III: Inputs, Fates, and Effects*, (National Academies Press, Washington, DC).
3. Atlas R & Bartha R (1992) Hydrocarbon Biodegradation and Oil Spill Bioremediation. *Advances in Microbial Ecology*, Advances in Microbial Ecology, ed Marshall KC (Springer US), Vol 12, pp 287-338.
4. Fuchs G (2008) Anaerobic metabolism of aromatic compounds. *Ann N Y Acad Sci* 1125:82-99.

5. Biegert T, Fuchs G, & Heider J (1996) Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem* 238(3):661-668.
6. Leuthner B & Heider J (2000) Anaerobic toluene catabolism of *Thauera aromatica*: the *bbs* operon codes for enzymes of beta oxidation of the intermediate benzylsuccinate. *J Bacteriol* 182(2):272-277.
7. Strijkstra A, *et al.* (2014) Anaerobic activation of *p*-cymene in denitrifying betaproteobacteria: methyl group hydroxylation versus addition to fumarate. *Appl Environ Microbiol* 80(24):7592-7603.
8. Harms G, Rabus R, & Widdel F (1999) Anaerobic oxidation of the aromatic plant hydrocarbon *p*-cymene by newly isolated denitrifying bacteria. *Arch Microbiol* 172(5):303-312.
9. Müller JA, Galushko AS, Kappler A, & Schink B (2001) Initiation of anaerobic degradation of *p*-cresol by formation of 4-hydroxybenzylsuccinate in *Desulfobacterium acetonicum*. *J Bacteriol* 183(2):752-757.
10. Wohlbrand L, *et al.* (2013) Complete genome, catabolic sub-proteomes and key-metabolites of *Desulfobacula toluolica* Tol2, a marine, aromatic compound-degrading, sulfate-reducing bacterium. *Environ Microbiol* 15(5):1334-1355.
11. Annweiler E, *et al.* (2000) Anaerobic degradation of 2-methylnaphthalene by a sulfate-reducing enrichment culture. *Appl Environ Microbiol* 66(12):5329-5333.
12. Selesi D, *et al.* (2010) Combined genomic and proteomic approaches identify gene clusters involved in anaerobic 2-methylnaphthalene degradation in the sulfate-reducing enrichment culture N47. *J Bacteriol* 192(1):295-306.
13. Callaghan AV, Wawrik B, Ni Chadhain SM, Young LY, & Zylstra GJ (2008) Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* 366(1):142-148.
14. Grundmann O, *et al.* (2008) Genes encoding the candidate enzyme for anaerobic activation of *n*-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* 10(2):376-385.
15. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.

16. O'Brien JR, *et al.* (2004) Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.
17. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA. How the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.
18. Becker A, *et al.* (1999) Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Biol* 6(10):969-975.
19. Leuthner B, *et al.* (1998) Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* 28(3):615-628.
20. Li L, *et al.* (2009) Subunit structure of benzylsuccinate synthase. *Biochemistry* 48(6):1284-1292.
21. Bhandare R, Calabro M, & Coschigano PW (2006) Site-directed mutagenesis of the *Thauera aromatica* strain T1 *tutE tutFDGH* gene cluster. *Biochem Biophys Res Commun* 346(3):992-998.
22. Coschigano PW (2002) Construction and characterization of insertion/deletion mutations of the *tutF*, *tutD*, and *tutG* genes of *Thauera aromatica* strain T1. *FEMS Microbiol Lett* 217(1):37-42.
23. Funk MA, Judd ET, Marsh EN, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci U S A* 111(28):10161-10166.
24. Vey JL, *et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.
25. Li L & Marsh EN (2006) Mechanism of benzylsuccinate synthase probed by substrate and isotope exchange. *J Am Chem Soc* 128(50):16056-16057.
26. Qiao C & Marsh EN (2005) Mechanism of benzylsuccinate synthase: stereochemistry of toluene addition to fumarate and maleate. *J Am Chem Soc* 127(24):8608-8609.
27. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.

28. Sievers F & Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079:105-116.
29. Verfurth K, Pierik AJ, Leutwein C, Zorn S, & Heider J (2004) Substrate specificities and electron paramagnetic resonance properties of benzylsuccinate synthases in anaerobic toluene and *m*-xylene metabolism. *Arch Microbiol* 181(2):155-162.
30. Li L & Marsh EN (2006) Deuterium isotope effects in the unusual addition of toluene to fumarate catalyzed by benzylsuccinate synthase. *Biochemistry* 45(46):13932-13938.
31. Himo F (2005) C–C bond formation and cleavage in radical enzymes, a theoretical perspective. *Biochim Biophys Acta* 1707(1):24-33.
32. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.
33. Bharadwaj VS, Vyas S, Villano SM, Maupin CM, & Dean AM (2015) Unravelling the impact of hydrocarbon structure on the fumarate addition mechanism—a gas-phase ab initio study. *Physical Chemistry Chemical Physics*.
34. Spormann AM & Widdel F (2000) Metabolism of alkylbenzenes, alkanes, and other hydrocarbons in anaerobic bacteria. *Biodegradation* 11(2-3):85-105.
35. Kniemeyer O, *et al.* (2007) Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* 449(7164):898-901.
36. Duncan KE, *et al.* (2009) Biocorrosive thermophilic microbial communities in Alaskan North Slope oil facilities. *Environ Sci Technol* 43(20):7977-7984.
37. Beasley KK & Nanny MA (2012) Potential energy surface for anaerobic oxidation of methane via fumarate addition. *Environ Sci Technol* 46(15):8244-8252.
38. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data. *Methods in Enzymology*, Vol 276.
39. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
40. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
41. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.

42. Eswar N, *et al.* (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5:Unit 5 6.

Chapter IV

Structural and biochemical characterization of choline trimethylamine-lyase suggests how a glyceryl radical enzyme binds and cleaves choline

The work described in this chapter was performed in collaboration with Prof. Emily P. Balskus (Harvard University). The biochemical experiments presented in the text were performed by Smaranda Bodea (née Craciun). The reagents and proteins used in this study were produced by Smaranda Bodea.

VI.A Summary

This chapter presents structures of the glycy radical enzyme choline trimethylamine-lyase (CutC) from *Desulfovibrio alaskensis* G20. CutC cleaves choline to acetaldehyde and trimethylamine through a net 1,2-elimination. The choline-bound structure of CutC suggests that electrostatics and CH–O interactions between the trimethylammonium moiety and polar or negatively charged protein residues are major contributors to choline binding and selectivity. There are some surprising differences between our X-ray structure and a previously reported homology model of CutC based on the structure of the closely related glycy radical enzyme glycerol dehydratase. There is excellent agreement in the lower half of the active site where the residues involved in the radical and deprotonation chemistry are located, consistent with conservation of chemical steps between these two systems. However, there is poor agreement in the upper half of the active site, where residues involved in binding and protonation of trimethylamine are found. In particular, variations in the β strand architecture position residues expected to be involved in catalysis away from the active site. A number of residues were identified in the CutC crystal structure as potentially important for catalysis, and site directed mutagenesis of these residues has allowed us to probe the mechanism of C–N bond cleavage. The relationship between glycy radical enzyme eliminases and the analogous adenosylcobalmin-dependent eliminases is discussed.

VI.B Introduction

The influence of the human gut microbiome on human health and the development of chronic diseases has recently become an area of much interest. The human gut is populated by a metabolically diverse group of bacteria that specialize in degradation of organic dietary compounds. One of these compounds is choline, an abundant metabolite used by humans for membrane biosynthesis, neurotransmission, and methionine biosynthesis and primarily acquired from the diet (1). Under strictly anaerobic conditions in the human gut, choline is degraded by diverse bacteria to acetaldehyde, which these organisms use as a source of energy and carbon, and trimethylamine (TMA) (Figure IV.1). A liver oxygenase converts TMA produced in the gut to trimethylamine-*N*-oxide (TMAO) (2), which is the subject of current study due to the correlation between TMAO levels and atherosclerosis (3-5).

Choline and its derivatives also play an important role in global nitrogen and carbon cycles. Within ruminants, TMA produced by choline-degrading bacteria causes off-flavor in milk or can be converted to methane by archaea, thus contributing to greenhouse gas production (6, 7). Once it is returned to the environment, TMA or TMAO derived from TMA can be degraded by methylotrophic bacteria. TMAO is also an important osmolyte and piezolyte for marine organisms (8) and is thought to be primarily derived from TMA produced by bacteria (9).

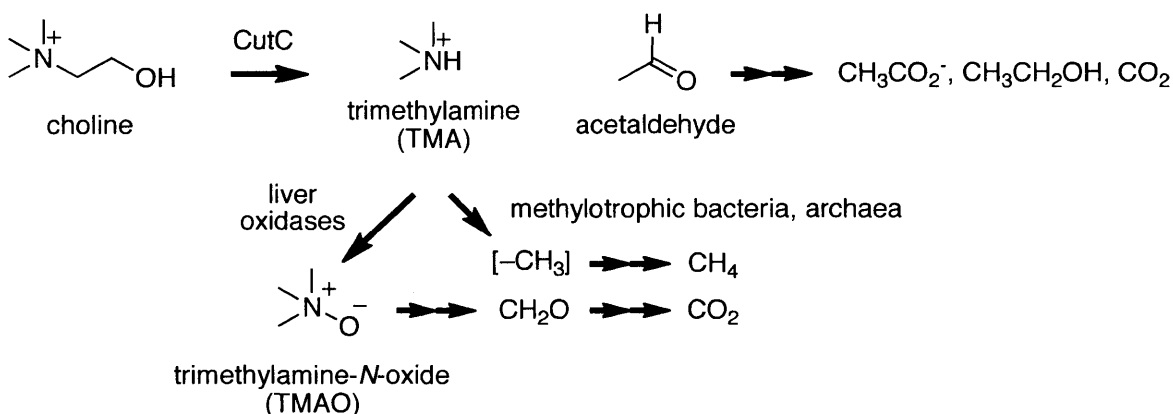


Figure IV.1. The role of choline trimethylamine-lyase (CutC) in initiating anaerobic choline metabolism. Choline is utilized by anaerobic bacteria after cleavage to trimethylamine (TMA) and acetaldehyde. Acetaldehyde generates ATP through fermentation or sulfate respiration or can be incorporated into cellular biomass. TMA and its oxidation product, TMA-N-oxide (TMAO), is a source of methyl groups for methylotrophic bacteria. Some archaea demethylate TMA and other methylamines, generating a methylated corrinoid intermediate that feeds into methanogenesis.

Although anaerobic choline degradation to TMA by bacteria was recognized over a century ago (10), the biochemical basis for this transformation was only recently elucidated (11). Choline cleavage requires a net 1,2-elimination, a difficult reaction that often requires the involvement of a radical species due to the high pKa of the alpha-carbon. In ethanolamine degradation by some bacteria, a similar reaction is catalyzed by the adenosylcobalamin-dependent (AdoCbl) ethanolamine ammonia-lyase (EAL), which cleaves the C–N bond of ethanolamine in a net 1,2-elimination to produce acetaldehyde and ammonia (12) (Figure IV.2A, IV.3A). A gene cluster containing many of the genes required for acetylaldehyde processing was identified in the choline-degrading organism *Desulfovibrio desulfuricans*. This provisional choline utilization (*cut*) operon contained two genes in place of EAL that coded for enzymes that could be involved in choline cleavage: CutC, a glycy radical enzyme (GRE), and CutD, an S-

adenosylmethionine (AdoMet) radical activating enzyme (AE). All GREs must be post-translationally modified by an AE, which abstracts a hydrogen atom from a glycine residue, installing a stable, backbone glycyly radical species. During GRE catalysis, this stable glycyly radical species is proposed to generate a transient thiyl radical as a radical initiator for diverse radical reactions (Figure IV.2B), including the well-characterized reaction of pyruvate formate-lyase (PFL), and the more recently characterized chemistry of both *p*-hydroxyphenylacetate decarboxylase (HPD) and benzylsuccinate synthase (BSS). Three other GREs are known to catalyze 1,2-eliminations: glycerol dehydratase (GDH) (13), which forms 3-hydroxypropionaldehyde during glycerol fermentation in *Clostridium butyricum*; propanediol dehydratase (PDH), which is highly similar to GDH in sequence and was discovered in a pathway for the degradation of the five-carbon sugar fucose in *Roseburia inulinovorans* (14); and class III or anaerobic ribonucleotide reductase, which catalyzes elimination of water from a ribonucleotide, forming a 2'-deoxy-3'-ketonucleotide intermediate (15, 16), which is then reduced by radical mechanism to yield the deoxynucleotide. CutC is ~37 % identical to GDH and thus represented a good candidate for a novel elimination reaction such as choline cleavage.

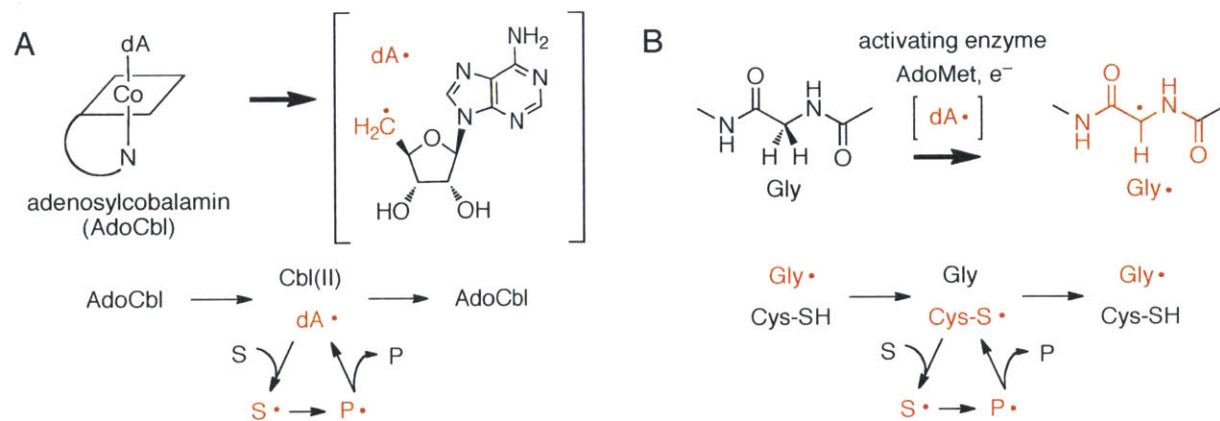


Figure IV.2. Similar radical generation strategies are found in AdoCbl and glycyly radical enzymes (GREs). (A) Adenosylcobalamin (AdoCbl) serves as a source of the 5'-deoxyadenosyl radical (5'-dA $^\bullet$), forming Co(II) cobalamin [Cbl(II)]. Substrates (S) are oxidized by 5'-dA $^\bullet$, rearranged in their radical form to a product radical (P $^\bullet$), and reoxidize dAH before leaving the active site. (B) 5'-dA $^\bullet$ is also used in the production of a stable glycyly radical within GREs by an S-adenosylmethionine (AdoMet) radical enzyme. The glycyly radical serves as a stable radical source, much like AdoCbl in A. A thiyl radical is proposed to serve as the active oxidant in all GREs and would be generated transiently within the active site. Substrate rearrangement occurs as in A, but may occur through different mechanisms.

The *cutC* and *cutD* genes were previously cloned and expressed heterologously in *Escherichia coli*, and the resulting proteins were shown to have choline TMA-lyase activity when incubated together with AdoMet and a reductant (17). A key question remaining in these studies is how choline binds and is activated for C–N cleavage. In the AdoCbl-dependent EAL and the related diol dehydratase, 1,2-elimination is proposed to occur through a 1,2-migration to yield a carbinolamine radical intermediate (Figure IV.3A) (12, 18, 19). CutC could potentially also perform such a migration reaction (Figure IV.3B), although it is not known how the bulky TMA moiety would be coordinated in the active site. An alternate proposal for CutC catalysis that does not invoke migration has been proposed (Figure IV.3C) (17). Deprotonation of the alcohol would permit cleavage of the C–N bond, forming a resonance stabilized aldehyde-based radical intermediate. This mechanistic proposal was inspired by density functional theory (DFT) calculations for GDH that indicate a base-catalyzed elimination is favorable in this system, and migration is associated with a very high energy barrier given the active site architecture of GDH (Figure IV.3D) (20). One explanation for a difference in mechanism between AdoCbl-dependent enzymes and GREs is the reactivity of the respective radical initiators: an AdoCbl-derived 5'-deoxyadenosyl radical (5'-dA•) is much more oxidizing, and thus more difficult to re-oxidize, than the proposed thiyl radical of GREs. Additionally, structures of both classes of enzymes have revealed that the binding sites are completely different, even when the substrates are similar or identical.

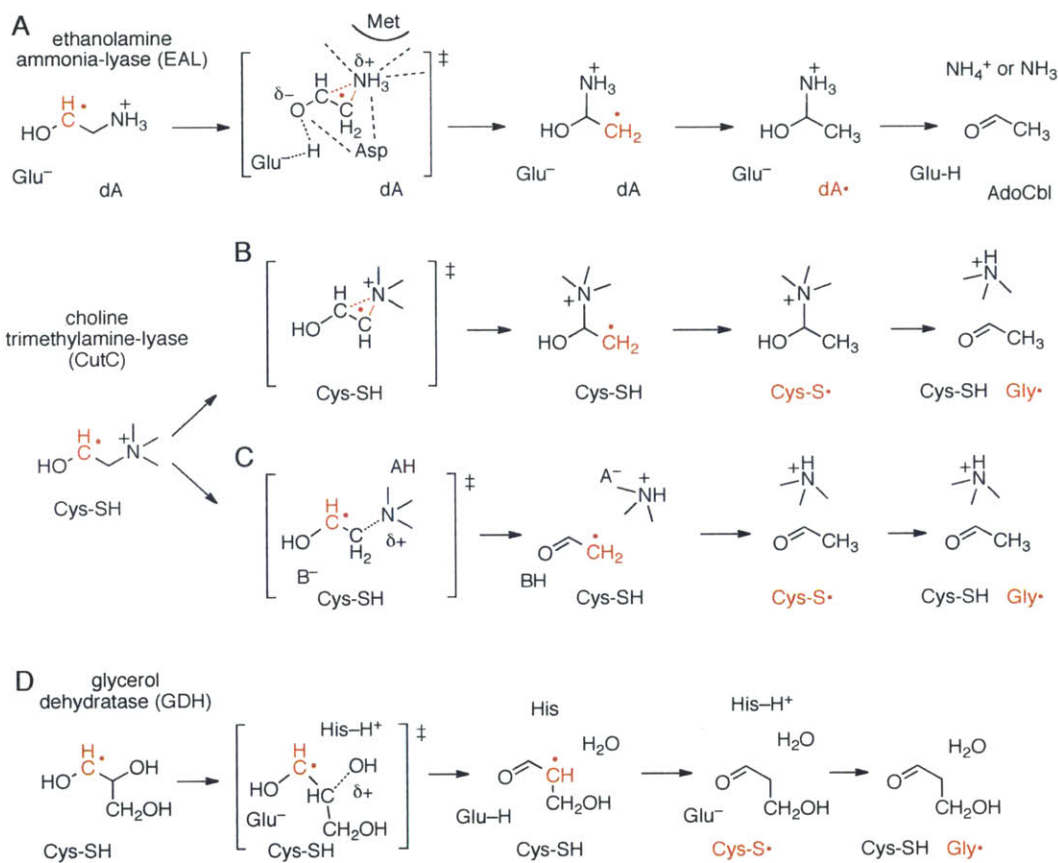


Figure IV.3. Proposed mechanisms of some radical 1,2-eliminases. (A) EAL is proposed to catalyze 1,2-migration of ammonia, through an associative, cyclic transition state (12, 18, 19). Elimination may be prevented by the absence of an appropriate proton donor; a methionine residue also sterically blocks loss of ammonia. A glutamate residue has been proposed to facilitate the migration by lowering the barrier height (19). The resulting carbinolamine radical is sufficiently oxidizing to regenerate 5'-dA. Elimination occurs after the radical is transferred back to the cofactor and may be facilitated by a general base (Glu). There are no apparent general acids within the active site. (B) By analogy to EAL, CutC could perform a 1,2-migration. (C) Alternately, a base-catalyzed elimination is possible. (D) GDH is proposed to catalyze 1,2-elimination. A proton donor (His-H⁺) and acceptor (Glu⁻) are present in the active site to facilitate this reaction.

To probe these mechanistic possibilities for CutC, we determined the structure of CutC from *Desulfovibrio alaskensis* G20 bound to choline. We also used a combination of additional structures and biochemical experiments to test the relative importance of residues identified by the structural analysis. We find that CutC contains an active site adapted to binding a small, charged substrate like choline. Comparison with the homologous GRE glycerol dehydratase reveals that a reorganized set of polar side chains coordinate the charged TMA moiety, whereas

the binding site for the C1 hydroxyl group is virtually unchanged. Mutagenesis of residues involved in binding the TMA moiety impaired—but did not abolish—catalysis, whereas mutants within a putative proton transfer network showed severe catalytic impairment in initial experiments. Our structure and mutagenesis results are most consistent with a mechanism involving base-catalyzed elimination of TMA, as proposed for GDH, rather than a 1,2-migration, which is the prevailing proposal for AdoCbl-dependent eliminases.

VI.C Results

Choline TMA-lyase adopts a conserved GRE fold.

We crystallized CutC from *D. alaskensis* G20 and solved the structure by molecular replacement with AdoCbl-independent glycerol dehydratase (GDH) (PDB ID 1R8W) (21) as the search model. The four molecules in the asymmetric unit form two dimers with a large shared surface area (Figure IV.4A). This tetrameric packing is similar to that seen in structures of *C. butyricum* GDH and *C. scatologenes* HPD, as well as *Archaeoglobus fulgidus* PFL2 (Figure IV.5), which is an enzyme of unknown function initially annotated as a PFL relative. A dimer/tetramer equilibrium has been proposed for PFL2 based on analytical ultracentrifugation experiments at 0.4–4 μM (22), and HPD is reported to form a hetero-octamer (a tetramer of the catalytic subunit bound to four accessory subunits) based on size-exclusion chromatography. All other characterized GREs are thought to be dimers (23). Previous size exclusion chromatography experiments indicated that CutC is a dimer even at the concentrations used here for crystallography (17), but the tetrameric assembly observed in the crystal structure may be relevant within bacterial microcompartments where CutC and some related GREs are thought to be localized (14, 24). Each CutC molecule adopts the canonical GRE fold: a ten-stranded β/α barrel with two loops containing conserved catalytic residues inserted into the center of the barrel (Figure IV.4B). A loop containing the catalytically essential glycine residue, the Gly loop, is found at the tip of the C-terminal glycy radical domain, a structural unit believed to undergo a conformational change during the installation of the glycy radical species by the activating enzyme (25, 26). An absolutely conserved glycine (G821 in CutC) is found at the tip of the Gly loop, fully in the interior of the barrel. Next to the Gly loop in the enzyme interior is the Cys loop, which resides at the tip of a pair of short, antiparallel β strands. The Cys loop contains an

absolutely conserved cysteine residue (C489 in CutC) and, depending on the GRE, other residues that contribute to catalysis (21, 27, 28). The active site cavity is found directly above the Cys loop, consistent with direct hydrogen atom abstraction from the substrate by a transient thiyl radical generated on C489 via G821. Alanine mutants of C489 and G821 were found to be completely inactive in the production of TMA in an end point assay (17), in agreement with these proposed roles in CutC.

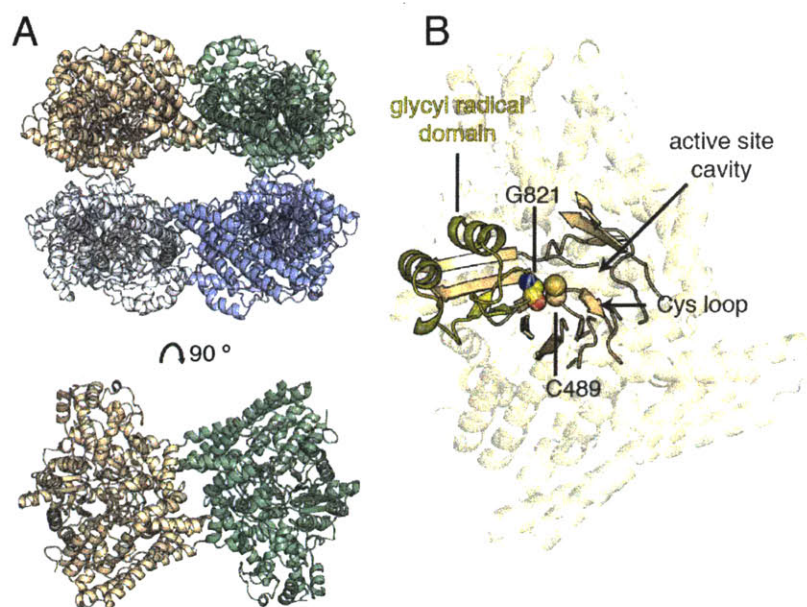


Figure IV.4. Overall structure of CutC from *D. alaskensis* G20. (A) CutC crystallizes with two dimers in the asymmetric unit. The tetrameric unit is similar to that of other GREs (Figure IV.5), and there are no apparent large differences between the four monomers. The dimeric unit (shown below rotated 90 °) is formed by a dimer interface conserved across all structurally characterized GREs other than class III RNR (see Chapter I.F). **(B)** The active site cavity is found at the center of the (β/α)₁₀ barrel (solid ribbon), above the Cys loop. Choline is present within the active site cavity but is not shown in this figure. Residues C489 (location of putative thiyl radical) and G821 (location of putative glycy radical) are shown in spheres. The glycy radical domain (yellow) harbors the active site glycy radical on the Gly loop in activated proteins.

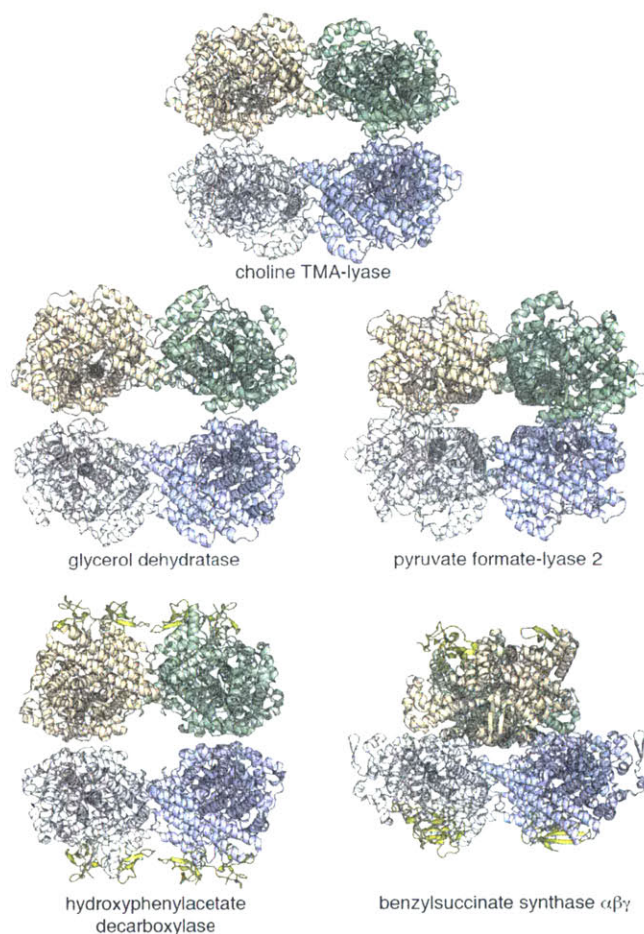


Figure IV.5. Tetrameric assemblies observed in the crystal lattices of GREs. For each GRE, four protomers are shown in roughly the same orientation. The structures used are 1R8W (21) for GDH, 2F3O (22) for PFL2, 2YAJ (29) for HPD (small subunit in yellow), and 4PKF (30) for BSS (both small subunits in yellow). For the BSS $\alpha\beta\gamma$ complex, the upper dimer is oriented 90° from where it is observed in the other GREs. No tetrameric assembly could be found for the BSS $\alpha\gamma$ complex (PDB ID 4PKC; not shown) or PFL (PDB 1CM5; not shown) (28). Only PFL2 and HPD have been observed to adopt a tetrameric (or heterooctameric for HPD) structure in solution (22, 29).

Choline interacts with polar residues within the CutC active site.

CutC was not supplemented with choline prior to crystallization, but density consistent with a bound choline molecule was present in the active site adjacent to C489, suggesting that choline co-purified with the protein. To ensure that substrate was bound at full occupancy in the crystal structure, crystals were soaked with excess choline (10 mM) prior to cryocooling. As in most other GREs, the structure of CutC with choline shows that residues from the Cys loop and

top face of the barrel contribute to substrate binding (Figure IV.6A). Hydrogen bonds from the Cys loop to the hydroxyl of choline, including the backbones amide of G488/C489 and the carboxylate of E491, hold C1 of choline in place adjacent to C489 S γ . In turn, the E491 carboxylate is held in place by hydrogen bonds from T502 and the Cys loop backbone. No residue in the Cys loop appears to play a role in selectivity for choline, consistent with the observation that the Cys loop sequence motif GCVEP is broadly conserved among diverse GREs, including GDH (see Figure IV.8).

The primary choline-selective interactions within the active site are with the phenolic oxygens of Y208 and Y506 and the carboxylate of D216 (Figure IV.6B). These oxygen atoms approach the methyl groups of choline closely, displaying C to O distances between 3.3 and 3.5 Å. Such interactions have been previously described as CH–O interactions or bonds (31) and are ubiquitous features in protein and nucleic acid structures that play an important, but underappreciated, role in stabilizing macromolecule structure at many levels (32, 33). In most contexts these weak interactions have been estimated to contribute to binding energy by only ~0.5-1.0 kcal/mol, comparable to a very weak hydrogen bond (33). In enzymes that bind quaternary amines and sulfonium groups, these CH–O interactions have been proposed to help determine specificity and facilitate catalysis (32, 34). In cases where a partially positive-charged C–H bond is present, estimates of the bond enthalpy are somewhat higher, ~1.2 kcal/mol (35). Within CutC, there are at least four close CH–O interactions (3.2-3.5 Å) and another three borderline interactions (3.5-3.7 Å) involving protein oxygen ligands and the TMA moiety. As choline is positively charged, there is likely a substantial electrostatic component to binding, as the first-shell residues D216 and E491 are negatively charged. Distances and a diagram of the protein-substrate interactions are given in Figure IV.6C. A single water molecule also contributes to binding through a CH–O interaction with choline and a hydrogen bond to Y506. The active site is thus highly polar and electron-donating, consistent with the diffuse, positive charge localized on the methyl groups of choline (36, 37). Choline also makes a face-on contact to the ring of F395 (C2 to ring center: 3.84 Å), which provides a cation- π interaction with the partial positive charge localized on choline C2 due to the quaternary ammonium. An aromatic residue is also conserved at this position in GDH, so this residue may have an alternate role within the active site (discussed below).

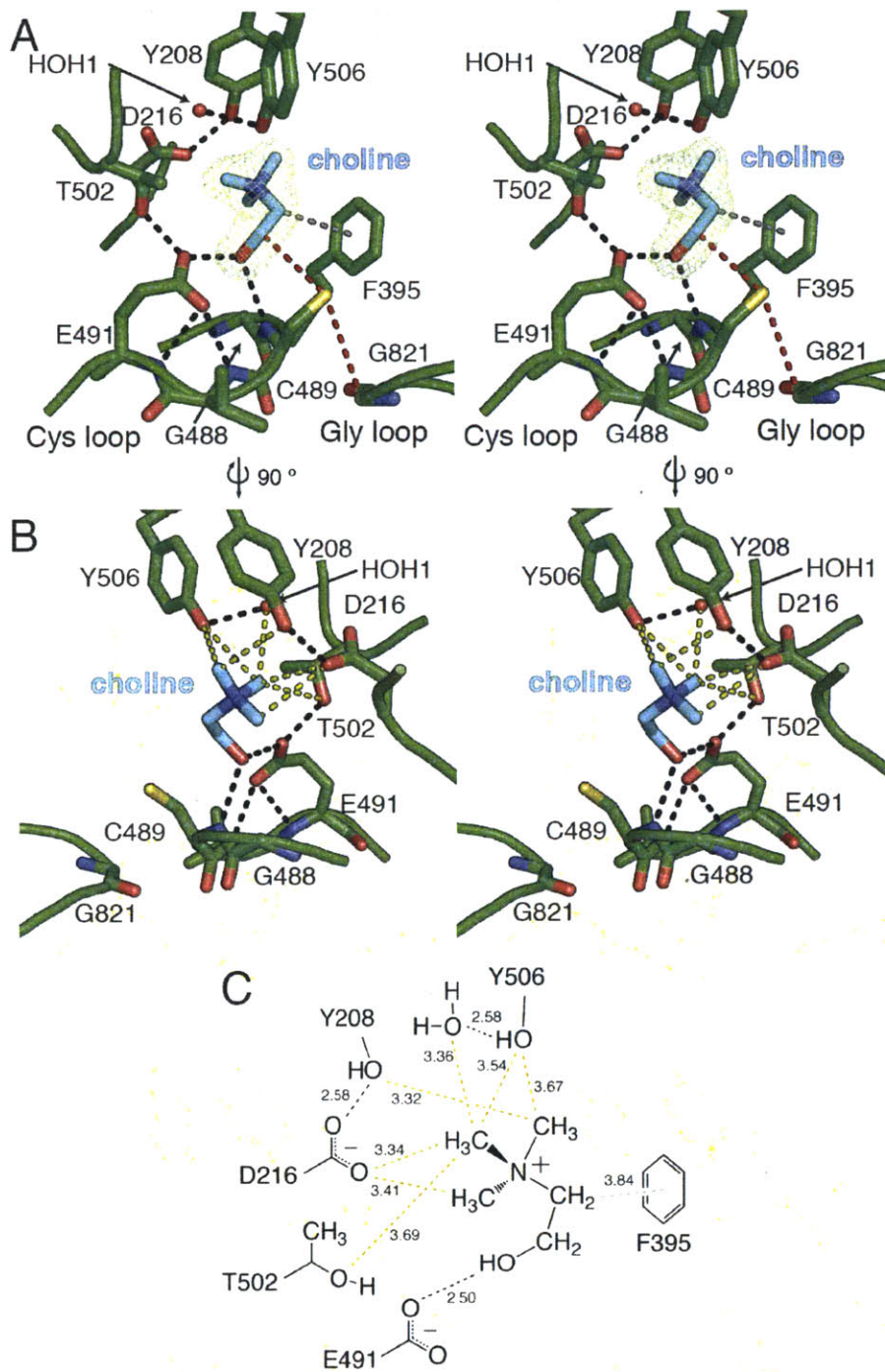


Figure IV.6. CutC binds substrate in a pocket above the Cys loop. (A) Stereoimage of CutC bound to its substrate, choline (cyan). The proposed hydrogen atom transfer pathway from G821 to C489 to C1 of choline is marked (red dashes). Hydrogen bonds (2.5-3.2 Å) are shown for residues within the active site (black dashes). **(B)** Stereoimage of the CH-O bonds (yellow dashes) present in the CutC-choline complex. **(C)** Diagram of protein and water interactions with the TMA moiety of choline with CH-O and hydrogen bond distances (Å). CH-O bonds are indicated for C to O distances of 3.8 Å or less (yellow). Hydrogen bonds (black) and presumed

cation- π interactions (gray) are shown between protein and substrate atoms. Distances (\AA) are given in the diagram. The maximum-likelihood-estimated coordinate error is 0.20 \AA .

Comparison of the CutC structure with its homology model reveals the difficulty in predicting active site features in GREs.

There are several important differences between the choline-bound structure and the homology model we previously constructed of CutC based on GDH (17) (Figure IV.8A-B). The active site architecture of CutC and GDH differ primarily at β 1 (containing Y208 and D216) and β 6 (containing T502), directly above the active site cavity and opposite to the Cys loop. β 1 is interrupted by a short helix in both CutC and GDH, connecting through β -sheet-like interactions with β 2 on one side of the helix and β 6 on the other. In the crystal structure of CutC, Y208 is found on this helix, pointing directly into the active site cavity. However, due to the low sequence similarity within β 1 between CutC and GDH and a single residue insertion upstream of β 1, the homology model is shifted out of register within this helix such that Y208 is facing out of the active site completely (Figure IV.8B). In contrast, the position of D216 on one of the sheet-like regions of β 1 is correctly modeled. Within β 6, T502 and Y506 are placed in essentially the correct position in the model, but the backbone conformation of the loop between these residues is different (Figure IV.8.C, with the result that S503 is not near the choline-binding site as predicted. It is notable that despite a relatively high identity overall between the two enzymes (37%), the strands that line the active site have very low sequence identity and structural similarity.

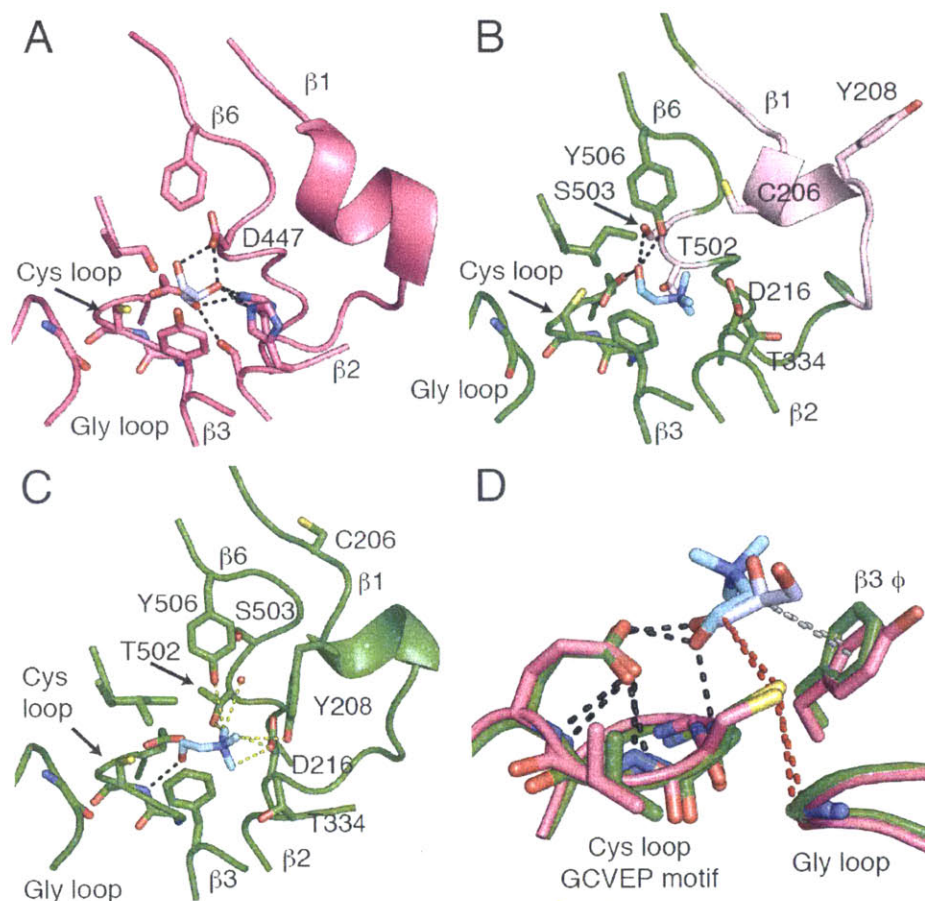


Figure IV.8. Comparison of CutC structure with GDH-based CutC homology model. The orientation of **A-C** is $\sim 90^\circ$ from that shown in Figure IV.6A. **(A)** GDH structure with bound glycerol (light blue). **(B)** Homology model of CutC created in Modeller using GDH as the starting model (17); choline (cyan) was docked using Schrödinger Suite 2012 (17). Residues that differ substantially from the observed positions in the crystal structure are highlighted (pink). **(C)** The crystal structure of CutC bound to choline (cyan) in the same orientation as in **B**. CH–O interactions (yellow dashes) and hydrogen bonds (black dashes) are shown as in Figure IV.6. **(D)** An overlay of CutC (green) and GDH (magenta) with residues involved in binding C1 and C2 (sticks); the orientation is similar to that in Figure IV.6. The Cys loop GCVEP motif is conserved between the two enzymes. An aromatic residue contributed from $\beta 3$ coordinates C2 (gray dashes). The proposed hydrogen transfer pathways (red dashes) between the active site Gly, Cys, and C1 are virtually identical in the two structures.

The docked orientation of the TMA moiety of choline in the homology model is correctly placed at the back of the active site (Figure IV.8B), but the hydroxyl of choline points away from the amide nitrogen of C489 in the homology model and instead forms hydrogen bonds with E491, Y506, and S503. In hindsight, the features responsible for binding C1, the C1-hydroxyl and C2 are essentially identical in the crystal structures of CutC and GDH (Figure

IV.8D), with the major changes in the active site of CutC due to the binding the trimethylammonium moiety at the rear of the active site. The positions of substrate C1 and C2 are likewise similar in these two members of the GRE family. Importantly, the abundance of CH–O bonds observed in the structure were not expected by the model, although D216 does make one close interaction. The absence of Y208 from the active site creates an open cavity, leaving one entire face of choline completely without protein contacts. Y506 makes only van der Waals contacts with the methyl groups of choline instead of close CH–O interactions.

Glycyl radical installation on wild type and mutant CutC proteins is variable.

One of the primary challenges in assaying GREs is generating the active form of the protein, which contains the glycyl radical cofactor. This step is typically performed in vitro using separately purified GRE and cognate AE, enzymatically produced and highly pure AdoMet, and a $1 e^-$ reductant. This reductant is an important variable that must be optimized for different enzyme systems, as it has been reported that different reductants yield different activation efficiencies (15), lead to uncoupling of AdoMet cleavage by the AE (38), and may also cause loss of the glycyl radical once it is formed (15, 39). Unfortunately, systematic optimization has not been reported for most GRE-AE systems. Flavodoxin in the semiquinone form is believed to be the in vivo reductant for *E. coli* RNR-AE (40, 41), PFL (42) and many other AdoMet radical enzymes (43, 44). Flavodoxin/flavodoxin reductase has been employed as the in vitro reductant for a number of AdoMet radical enzymes (45-49). Non-physiological reductants are often easier to work with, especially for EPR experiments, and have been reported to support higher activity than the presumed in vivo systems, likely due to the lower redox potential of most non-physiological reductants (40, 48, 50). Dithionite is a common choice, but it is unstable in solution, is known to cause non-productive AdoMet cleavage in many systems, including GRE-AEs (38), and has also been observed to react with a product-based radical in the AdoMet radical enzyme spore photoproduct lyase (51). Alternate non-physiological reductants, including Ti(III) citrate or a photoreductant such as deazaflavin or rose Bengal (15, 40), are also available.

For systems in which the radical content of GREs has been quantified by EPR, in vitro activation typically yields no more than 0.5 radicals per polypeptide (38, 41, 52), and it has been suggested for PFL and class III RNR that half-of-the-sites reactivity leads to a single radical site per dimer (53). Although in vitro activation results consistently yield no more than 50 %

activation (as assessed by SDS/PAGE of the oxygen cleaved protein), it has not been conclusively shown that the activated and unactivated polypeptides are distributed as a 1:1 complex. Analysis of the protein cleavage products of BSS (54), class III RNR (55), and PFL (56) after exposure of cells to oxygen suggests that 50 % activation is also the case in vivo. The molecular basis of this apparent half-of-the-sites activation is not clear as the glycy radical domains within the dimer are >70 Å away from each other. For HPD, which is proposed to form a tetramer of $\alpha\beta$ units, it has been suggested that a maximum of 0.25 radicals per polypeptide are produced (57), although this result has not been confirmed in vivo. Likewise, it was suggested that this lower activation reflected a single radical per $\alpha\beta$ tetramer, but there is no direct evidence for this claim. In vitro activation has been reported for GDH, but the glycy radical content was not quantified. Additionally, there remains serious controversy over the activity of GDH-AE, as this enzyme is reported to produce 2-aminobutyrate instead of 5'-dA as the product of homolytic AdoMet cleavage (58), despite no substantive difference in the sequence of the AE from related enzymes known to produce 5'-dA. No in vitro activation has yet been reported for BSS.

To determine which residues near the active site are important for catalysis, we have employed site-directed mutagenesis to study active site residues that are involved in choline binding or that play a role in catalysis (Table IV.2). Initial in vitro characterization was performed based on the homology model prior to the determination of the crystal structure of CutC (17). An overnight (16 h) end point assay in which TMA is detected directly by liquid chromatography-mass spectrometry (LC-MS) was used to assess whether the mutants were capable of any TMA production. C489 and G821, the Cys and Gly loop residues thought to be essential for radical catalysis and storage in all GREs, were previously shown to be completely inactive by this assay (17). The CutC mutants E491Q, D216N, and F395L were also incapable of producing TMA in the end point assay, but the source of the inactivity was not established previously.

To assess whether the inactivity observed for CutC mutants is due to a defect in activation, we chose to determine the level of activation of wild-type CutC and CutC mutants by using electron paramagnetic resonance (EPR) spectroscopy to perform spin quantitation of the glycy radical signal (Table IV.2). The activation reaction was initially optimized using a coupled, spectrophotometric assay for acetaldehyde, described below for steady state kinetics. Parameters optimized for the activation reaction include ratio of CutC to CutD, incubation time,

reductant, reductant concentration, and including an additional gel filtration step in the CutC purification protocol. After extensive optimization, we observe at most ~16 % activated polypeptide for wild-type CutC, which is well below the 50 % observed in PFL, but similar to the activation yields reported for HPD (38, 57). Given that none of our CutC mutations affect residues of the glycy radical domain, we were surprised to find substantial decreases in the glycy radical content for a number of the mutant proteins. CutC-D216N, CutC-Y506F and CutC-T502A are not activated to any measurable extent. In contrast, CutC-E491Q, CutC-E491A, CutC-Y208F, CutC-T502S, and the CutC-Y208/Y506F can be activated, albeit to a lesser extent than wild type. Surprisingly, the CutC-F395L mutant yielded the highest radical content, slightly greater than wild type. F395 is adjacent to the Gly loop, suggesting that some change in interaction between these residues could lead to enhanced activation (Figure IV.6A). Two possibilities can be considered for the reduced CutC activation of several of the active site mutants: 1) the mutation disrupts the glycy radical domain such that the glycy radical cannot be generated or 2) the radical is generated normally but is reduced at a vastly elevated rate. We are currently attempting to resolve these questions by assaying for 5'-dA formation concurrent with glycy radical formation.

In order to characterize the kinetics of choline cleavage by activated CutC, we used a combination of techniques: GC-MS to monitor TMA production and a coupled, spectrophotometric assay to detect acetaldehyde production. The latter assays were conducted in 96 well plates with NADH consumption by yeast aldehyde dehydrogenase monitored by a plate reader. The spectrophotometric assays contained 200 μ M NADH, 25 mM Tris HCl buffer pH 8, 50 mM NaCl, and varying volumes of diluted activation mixture, based on the necessary concentration of protein. The choline concentration ranged from 0-10 mM for wild type CutC with low K_M , up to 0-150 mM for mutants with higher K_M . Previous controls using the same assay setup for which no activity was observed included aerobic incubation or omission of AdoMet, dithionite, choline, CutD, CutC or choline (17).

Kinetic assays exhibit linear behavior over at least 5 min for wild-type CutC (~21,000 turnovers per polypeptide or ~120,000 turnovers per radical), suggesting that once the glycy radical is formed within CutC it is unlikely to be reduced substantially within this time frame. The activation mix is diluted 2000-fold to a final wild-type CutC concentration of 5 nM, so reactivation of the protein by the remaining activating enzyme is highly unlikely. Similar linear

fits are obtained for CutC mutants, although those with vastly impaired activity are not diluted prior to the assay, so it is difficult to be confident that they are not re-activated. Nevertheless, activation kinetics appear to be slow (activity peaks at 30-60 min), so continual reactivation is unlikely to explain the stable activity. These results suggest that the mutants retain the glycy radical through many rounds of turnover, despite impaired catalytic function, and verification of this result is ongoing.

protein	radicals per polypeptide	detectable activity (TMA)	K_M (mM)	k_{cat} (s^{-1})	glycyl-radical-normalized k_{cat} (s^{-1})
wild type	8.5	yes	0.33	70 ± 3	820
E491Q	2.4	no	ND	ND	ND
E491A	6.0	no	ND	ND	ND
F395L	10.5	yes	1.5 ± 0.5	0.022 ± 0.002	0.2
Y208F	5.0	yes	2.5 ± 0.1	3.52 ± 0.07	70
Y506F	<1	yes	1.3 ± 0.1	>1	>140
Y208F/ Y506F	2.4	yes	18.6 ± 1.5	0.58 ± 0.02	25
T502A	<1	yes	50.7	>0.05	>5
T502S	7.0	yes	3.9 ± 0.6	65 ± 3	920

Table IV.2. Activation and kinetic parameters of CutC mutants. Activation was measured by double integration of the EPR signal from each mutant after activation by CutD (see methods). The values reported with standard deviation are the average of three replicates. Activity for the mutants was initially detected by analysis of TMA production by an overnight end point assay for TMA. Mutants with detectable activity were kinetically characterized by a coupled assay as described in the methods. The un-normalized k_{cat} is given based on the total enzyme concentration. Additionally, a normalized k_{cat} based on the radical concentration was estimated by dividing by the fraction of activated protein.

Mutations of tyrosines involved in CH–O interactions yield enzyme that is still active.

Based on the crystal structure of CutC, the mutants Y208F and Y506F were tested to determine if the proposed CH–O bonds are indeed important for binding of or activity on choline (Table IV.2). CutC-Y208F is activated by CutD to an extent comparable to wild-type CutC, yet kinetic analysis revealed a ~20-fold decrease in k_{cat} and ~150-fold decrease in k_{cat}/K_M . CutC-Y506F showed a serious activation defect, with no observable radical signal. Nonetheless, the enzyme was capable of producing TMA in the end point assay described above. Kinetic analysis revealed a lower bound on the radical-normalized k_{cat} of $140 s^{-1}$ for the mutant based on an

estimated EPR detection limit of ~1%, suggesting that activity could be reduced, but the impact on activation is larger. The combined Y208F/Y506F mutant had substantially diminished k_{cat} and K_M values, similar to the Y208F single mutant protein (Table IV.2).

To further probe these mutant enzymes, we crystallized CutC-Y208F after incubation with 10 mM choline. A mostly unperturbed active site is observed in the structure, with choline bound similarly to wild-type CutC (Figure IV.9A). A water molecule is found in two of four of the molecules in the asymmetric unit in place of the phenolic hydroxyl but close contacts with adjacent residues yield lower than 100% occupancy for this water. We are pursuing a structure of CutC-Y506F as well as isothermal titration calorimetry (ITC) experiments in order to directly compare the impact of these two mutations on choline binding.

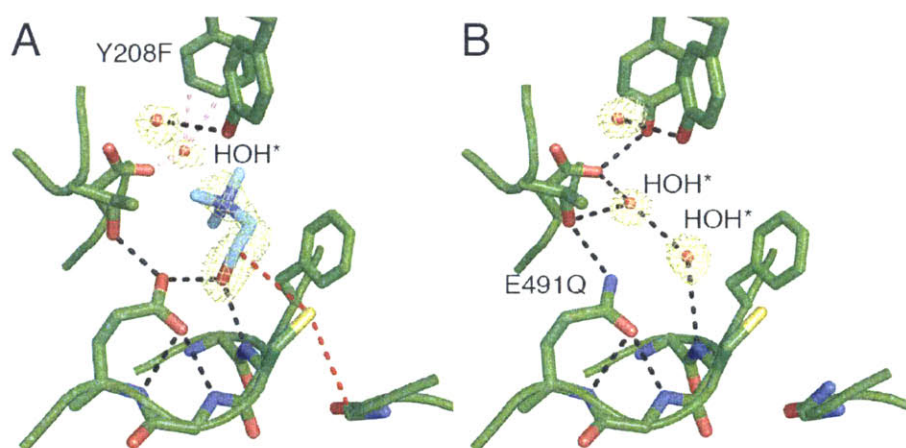


Figure IV.9. Structures of mutant CutC enzymes show bigger effect on choline binding for E491Q than for Y208F. Crystals were grown in the presence of 10 mM choline. $F_o - F_c$ omit density is contoured at 3σ ; choline and all water molecules were omitted prior to calculation of the map, **(A)** CutC-Y208F (1.95-Å resolution) introduces a new water molecule in the rear of the active site (HOH*). Hydrogen bonds (black dashes), steric clashes (pink dashes), and hydrogen-atom transfer pathways (red dashes) are indicated. **(B)** The E491Q mutant (1.70-Å resolution) has no density for choline. Instead, two new water molecules are bound in the active site (HOH*). The orientation of the side chain amide is deduced based on the contacts to hydrogen bond donors within the Cys loop.

Mutation of residues in conserved GRE motifs suggests conservation of function as well as sequence.

Figure IV.8D highlights a conserved Cys loop GCVEP motif and a conserved aromatic residue off of $\beta 3$ (F395 in CutC). As described above, CutC-F395L is unusual among all eight

CutC single mutant proteins in that its glycy radical content is higher than wild type. However, kinetic parameters, which were determined using a coupled assay observing acetaldehyde reduction, showed that CutC-F395L is indeed active, although both K_M and k_{cat} are dramatically impaired (Table IV.2). We are pursuing a structure to determine if choline is bound correctly in the active site and additional biochemical experiments to characterize the catalytic defect of this mutant.

The mechanistic proposal in Figure IV.3C requires an acceptor for the choline hydroxyl proton, and E491 of the conserved GCVEP motif appears positioned to play the role of proton acceptor in CutC, as in GDH (Figure IV.8D) and class I and II RNRs (discussed below). Mutations of E491 would therefore be expected to have a large affect on catalysis, and we find that both CutC-E491Q and CutC-E491A are completely inactive in an overnight end point assay (see above and methods) (Table IV.2). To further investigate the source of the inactivity, we determined the crystal structure of CutC-E491Q. As before, we pre-incubated protein with 10 mM choline prior to setting up crystal trays and 10 mM choline was included in the cryoprotectant solution. The resulting structure revealed that mutant and wild-type enzymes were virtually identical in structure [root mean square deviation (RMSD) of 0.18 Å] including the arrangement of residues in the active site (RMSD of 0.09 for Cys loop residues). Despite pre-incubation of the protein with choline, no density whatsoever was observed for choline within the active site (Figure IV.9B). Instead, two new water molecules are bound, one replacing the hydroxyl oxygen of choline, and the other making contacts to D216 and T502. Analysis of predicted hydrogen orientations suggests that in wild-type CutC, the alcohol acts as a hydrogen bond donor to E491 and acceptor to the backbone nitrogen of the Gly488/Cys489 peptide bond (Figure IV.10A). There are no other hydrogen bond donors or acceptors near choline. As there are no changes in the overall structure of this mutant, we take the absence of density for choline in the active site as an indication that CutC-E419Q has a dramatically reduced affinity for choline. In the E419Q mutant structure, the amide side chain of glutamine is positioned analogously to the carboxylate of glutamate, such that it can hydrogen bond to the same backbone amide nitrogen of the Cys loop. This orientation of glutamine means that the side chain amide carbonyl is facing away from the choline-binding site where it can serve as a hydrogen-bond acceptor of the backbone amide, and the side chain amide nitrogen, which is a hydrogen bond donor, is now pointing toward the choline-binding site where it hydrogen-bonds with a

water molecule rather than choline (Figure IV.10B). Thus, in the E491Q structure, there is no hydrogen bond acceptor for the choline hydroxyl, and the binding of water is favored over choline. E491 also interacts with T502, and in the E491Q structure, the hydrogen bond is reversed such that T502 is now the acceptor and E491Q is the donor. This forces the T502 O–H bond to rotate and face toward a water molecule now present in the active site. This rotation also shortens the distance between the T502 hydroxyl group and the D216 carboxylate to ~ 3 Å.

In addition to the impairment of choline binding, mutation of E419 would also be expected to impair catalysis by eliminating the proton acceptor for the choline hydroxyl group. We are attempting experiments with different mutants at this position to assess the role of a proton acceptor at this position.

A putative proton transfer network in the active site of CutC.

With the above data supporting a role for E491 as the catalytic base, multiple turnovers will require that E491 be deprotonated after each round of catalysis. However, the active site of CutC does not appear to be set up to allow transfer of protons from the exterior of the protein as there is no direct pathway access between the active site and solvent in any of our crystal structures. Additionally, TMA is expected to be protonated prior to release from the active site and it is tempting to consider if the same proton removed from the choline substrate could be used to protonate the TMA product. E491 is part of a hydrogen-bonding network (E491-T502---D216-Y208), which could act as a proton transfer network as well, albeit the T502 to D216 distance is too long (4.0-4.6 Å) in our current structures of wild-type CutC to support proton transfer (Figure IV.10A). The proximal residue to E491 in this hydrogen-bonding network is T502, prompting us to investigate the importance of T502 to the CutC reaction. T502 is absolutely conserved in CutC-like enzymes from a diverse group of bacteria, despite no direct contact between the γ -hydroxyl or γ -methyl to choline in the wild type CutC crystal structure. Initial TMA-production experiments indicated that the T502A and T502S mutants were both active (17). CutC-T502S appears to have activity comparable to wild type, but with a ten-fold elevated K_M .

CutC-T502A is severely deficient in glycy radical formation, making it difficult to assess the true activity of this mutant protein. Because of the inability to activate the protein, it is impossible to evaluate the effect of the mutation on the enzyme activity directly. The K_M for

CutC-T502A is substantially perturbed from that of wild-type CutC—the highest of all the mutants for which this parameter could be determined—suggesting that there is indeed a defect in catalytic activity for this mutant.

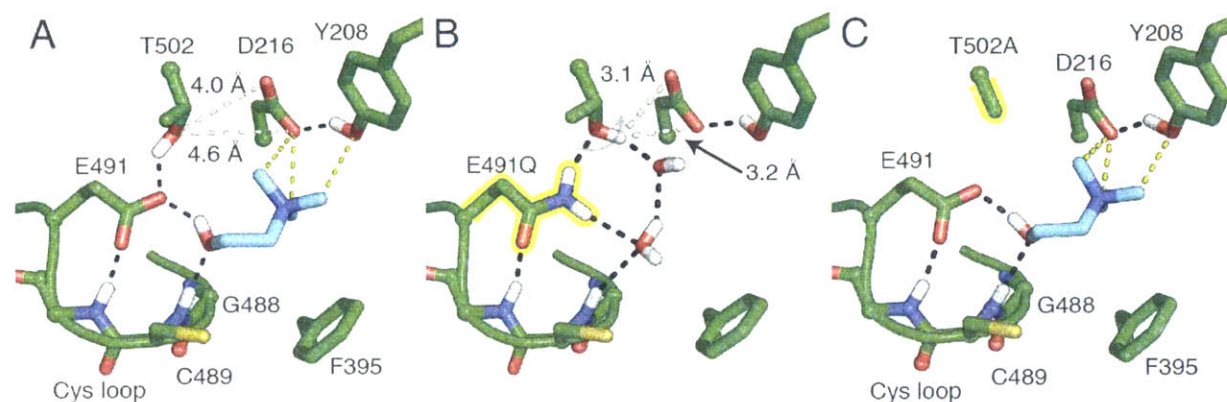


Figure IV.10. Putative hydrogen bonding networks in wild-type and mutant CutC enzymes. The orientation of this figure is $\sim 90^\circ$ around the y-axis from that in Figure IV.6. Y506 and the Gly loop are not shown for clarity. In all structures the pH of the crystallization was 7.0. **(A)** In wild-type CutC, two networks of hydrogen bonds are highlighted (black dashes). A proposed connection (white dashes) between T502-D216 would be possible if either of these residues were to change conformation. CH–O bonds to Y208 and D216 are shown as in Figure IV.6. Polar hydrogen atoms are shown within the networks based on inferred position given the available donor and acceptor groups. Hydrogens were not modeled explicitly in the crystallographic refinement, but were added manually to the final model. **(B)** In CutC-E491Q, in addition to the absence of choline, the inferred arrangement of hydrogen atoms must rearrange to accommodate the two hydrogens on the amide side chain. **(C)** In CutC-T502A, no connection is available between E491 and D216.

To further address the role of T502 in CutC catalysis, we crystallized both mutants after incubation with 10 mM choline in buffer at pH 8.0. Within the resulting structures, both mutants appear to bind choline identically to wild-type CutC (Figure IV.11B-C). The positions of the side chains of E491 in both mutants and the T502 hydroxyl group in T502S were unchanged, suggesting that T502 is not vital for the positioning of E491 or the hydroxyl of choline. Examination of the predicted hydrogen-bonding network within CutC-T502A reveals that there were no changes in interactions around the C1 hydroxyl or the TMA moiety (Figure IV.10C); the only difference is that E491 can no longer hydrogen bond to the side chain of T502, as it is no longer present. The fact that the T502A mutation does not influence the positions of other residues in the active site, apparently binds choline at full occupancy, and yet shows a greatly elevated K_M value supports a role for T502 in some catalytic step other than binding.

If T502 is part of a proton transfer pathway that regenerates the deprotonated state of E491 by transferring the proton to product TMA, then another residue must also be involved as the T502 to TMA distance (5.0 Å) and angle (104.5° choline C2–N bond to T502 γ -OH) are not appropriate for a direct transfer. As mentioned above, D216 is close (4.0-4.6 Å) to T502, but not close enough in its current position for hydrogen bonding. Interestingly, modeling of hydrogens in the structure of CutC-E491Q (Figure IV.10B) suggests that small structural changes can decrease the distance between T502 and D216 by 1 Å or more, consistent with the possibility of a proton transfer between these two residues. It is important to note that we do not expect T502 to be deprotonated, but rather to serve as a proton shuttle between E491 and D216. Experiments with mutants of D216 are ongoing and will hopefully shed more light on the importance of proton transfer within the CutC active site.

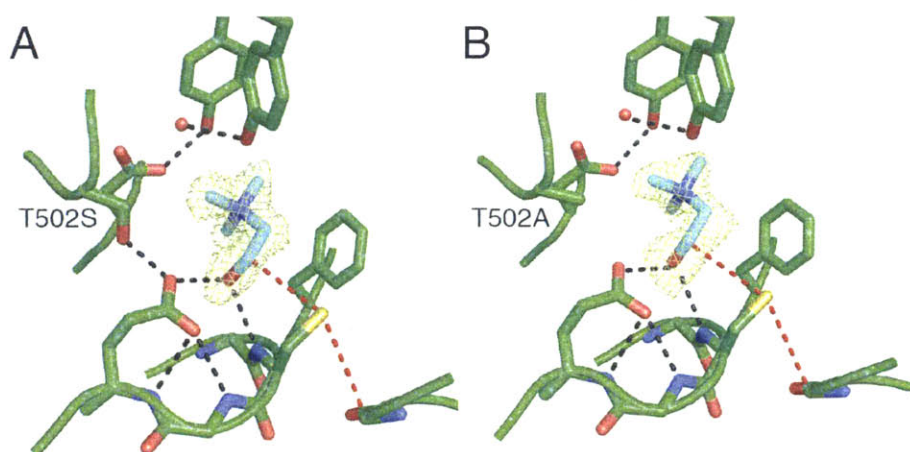


Figure IV.11. T502 mutants are able to bind choline. (A) T502S at 1.85-Å resolution; **(B)** T502A at 1.85-Å resolution. Crystals were grown in the presence of 10 mM choline. $F_o - F_c$ omit density is contoured at 3σ ; choline was omitted in the map calculation. Hydrogen bonds (black dashes) and hydrogen atom transfer steps (red dashes) are shown.

IV.D Discussion

Choline and its derivatives have important impacts on human health and microbial growth and energy conservation. The discovery of a long sought after anaerobic microbial pathway for choline degradation initiated by a glycyl radical enzyme was enabled by comparisons to the well-studied AdoCbl enzyme ethanolamine ammonia-lyase and GREs such as glycerol dehydratase. The structures and biochemical experiments presented here reveal

important differences between CutC and both of these enzymes as well as broader lessons about the challenges radical enzymes must overcome to be effective catalysts.

One of the interesting features of the CutC active site is how this enzyme recognizes the TMA moiety of choline. In a number of structurally characterized choline-binding proteins, the TMA moiety of choline interacts with aromatic amino acids through apparent cation- π interactions (Figure IV.12A). In contrast, CutC only has a single aromatic residue in the active site: F395, which makes a contact to C2. However, this residue is also present in GDH and thus appears to be contributing to catalysis in a different way in this system. Instead, CutC uses polar residues to make CH-O bonds to the TMA-moiety (Table IV.1). To establish whether other proteins that bind TMA-like molecules do the same, we inspected the active sites of nine proteins and enzymes shown in Figure IV.12, which bind choline, glycine betaine, carnitine, or trimethyllysine, and find that CH-O bonds are quite common. Typical quaternary amine binding sites contained two to four CH-O bonds, complemented by cation- π /van der Waals interactions from the tryptophan, phenylalanine, and tyrosine residues. Relative to choline-binding proteins CutC contains a surprising number of CH-O bonds (six with distances less than 3.5 Å), perhaps reflective of the unique chemistry that occurs at the TMA moiety in this system.

(figure next page)

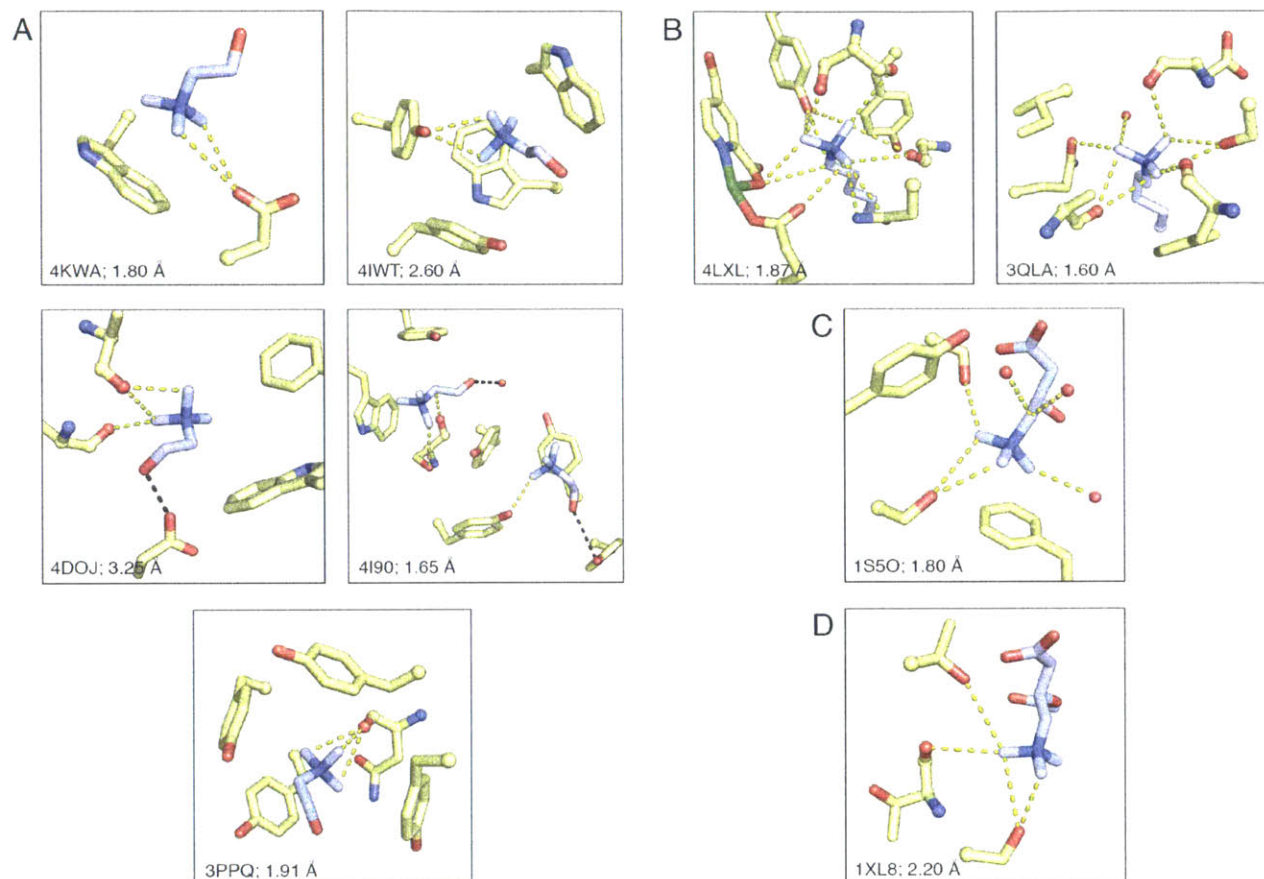


Figure IV.12. CH–O bonds are common in quaternary-ammonium-binding proteins. Proteins show a mixture of CH–O and cation- π interactions when binding **(A)** choline [(59-62) and unpublished (4KWA)], **(B)** trimethyllysine [(63) and unpublished (4LXL)], **(C)** carnitine (64), **(D)** octanoylcarnitine (65). PDB IDs and reported resolution are listed within each panel. All contacts closer than 3.7 Å were taken to indicate CH–O bonding. The average distance observed was 3.35 ± 0.2 Å. The median number of CH–O interactions per structure was three, with a large variation.

Only two members of GRE eliminase family, GDH and CutC, have a structure with a bound substrate. It is apparent from comparison of these enzymes that specificity is conferred primarily by β barrel strands β_1 , β_2 , and β_6 , which make the top surface of the active site and provide side chains interactions to specific substrates. Despite the high identity between these enzymes (37%), homology modeling missed a number of key active site features, a problem that may turn out to be common for GRE enzymes. We suspect that the difficulty of modeling the CutC active site is due to a combination of factors: challenges associated with the divergent β strand architecture (presence of β strand bulges, specifically) within the core of the protein, a lack of parameters for modeling the CH–O bonds that contribute to binding the TMA moiety,

and the inability to restrict the solutions found by docking programs to ones that make chemical sense. The latter problem could be overcome for these eliminases by including a substrate pre-placed in the active site in the correct orientation for radical abstraction and elimination. Hydrogen atom abstraction takes place in essentially the same place in CutC and GDH (Figure IV.8D), despite other large differences in how substrate is bound. The high degree of flexibility in active site strands in the GRE enzyme class is a more difficult modeling problem to solve. In addition to modeling GREs with known function but unknown structure, previous bioinformatics efforts (27) have identified a large number of GDH- and PFL-like genes that are 30-45 % identical to known enzymes, but are clearly not GDHs or PFLs based on the replacement of conserved residues within the Cys loop. These enzymes may contribute to new metabolic pathways that have yet to be characterized. Thus, tools that will help predict the nature of the substrate based on active site sequence are also valuable.

In terms of enzyme mechanism, a key question we sought to address is how the C–N bond of choline is cleaved by CutC. Two precedents for mechanism were available: the well-studied 1,2-migration catalyzed by the AdoCbl enzyme ethanolamine ammonia-lyase is perhaps the closest chemical analog of CutC as it involves C–N bond cleavage; however, the GRE glycerol dehydratase is much more similar in terms of overall structure and active site arrangement, despite the β strand variations mentioned above. Density functional theory (DFT) calculations have shown a lower energetic penalty for base-catalyzed elimination in GDH relative to migration (18, 20). Experimentally, differentiating between these two potential mechanisms is quite challenging in GREs. However, our combined crystallographic and biochemical data prompt us to favor a base-catalyzed elimination model for CutC catalysis similar to that proposed for GDH. In the mechanistic proposal shown in Figure IV.13, we predict based on our structures that the choline C1 radical is formed by abstraction of the pro-*S* hydrogen by a C489 thiyl radical (step I). Heterolytic C–N bond cleavage is catalyzed by deprotonation of C1-OH by E491 and protonation of the departing TMA group by D216 (step II-IV). From our structures, it is apparent that choline/E491/T502 and D216/Y208 form two hydrogen bond networks that may be able to communicate to enable proton transfer between the choline C1 hydroxyl and TMA (Figure IV.10A). This proton shuffling could be important in CutC because of the closed nature of the active site, which likely limits exchange between bulk solvent and polar active site residues. Using the proton removed from the C1-OH group of choline to

protonate the TMA group ensures that protonation states within the active site are reset after every round of turnover. It is possible, however, that in the active form of the enzyme, protons can freely move between the active site and the exterior solvent. Computational studies may eventually help determine the importance of proton movement within the active site of CutC.

In the choline-bound structure, there is no direct connection between D216 and T502, but these residues are only ~ 4.0 Å apart. Simple rotation of T502 around χ_1 by 10 - 15° would allow formation of a new hydrogen bond between T502 and D216. This shift of D216 toward T502, which could be prompted by movement of choline in response to radical abstraction or deprotonation of C1, would allow T502 to mediate proton transfer between E491 and D216. The net result would be movement of the proton from E491 to D216. In agreement with the proposed importance of the connection between T502 and D216, we find that the mutants T502A and D216N are impaired in choline binding, although we are still working to verify that this is a catalytic defect and not related to activation of the proteins.

The positioning of D216-Y208 directly behind the C–N bond suggests that proton transfer to TMA could occur concurrently with C–N bond cleavage. If this is the case, protonation of D216 likely precedes C–N bond cleavage, suggesting that removal of the proton from the choline C1 hydroxyl by E491 to form a transient ketyl radical zwitterion could be a prerequisite to choline cleavage (Figure IV.13, step III). Ketyl radical anions have been proposed elsewhere for analogous radical dehydratases (66-68) and, in the case of CutC, would likely be stabilized by the adjacent quaternary ammonium. The pKa of the choline hydroxyl [~ 13.9 (69)] is already lower than a typical alcohol, and the pKa of α -hydroxy radicals is typically 4-8 pH units lower than the corresponding alcohol (70), suggesting that proton transfer to E491 to form a ketyl radical is quite reasonable. In class Ia RNR, this mutant permits dehydration of the ribonucleotide, albeit at substantially reduced rates (71, 72), but also causes off-pathway reactions resulting in enzyme inactivation due to the inability to reduce the dehydrated product. Unfortunately, choline does not appear to bind at all to the E491Q mutant of CutC due to loss of an important hydrogen bond, so it is difficult to assess the impact of removal of the general base on the enzyme-catalyzed reaction, as has been done in the case of class Ia RNR (72-74).

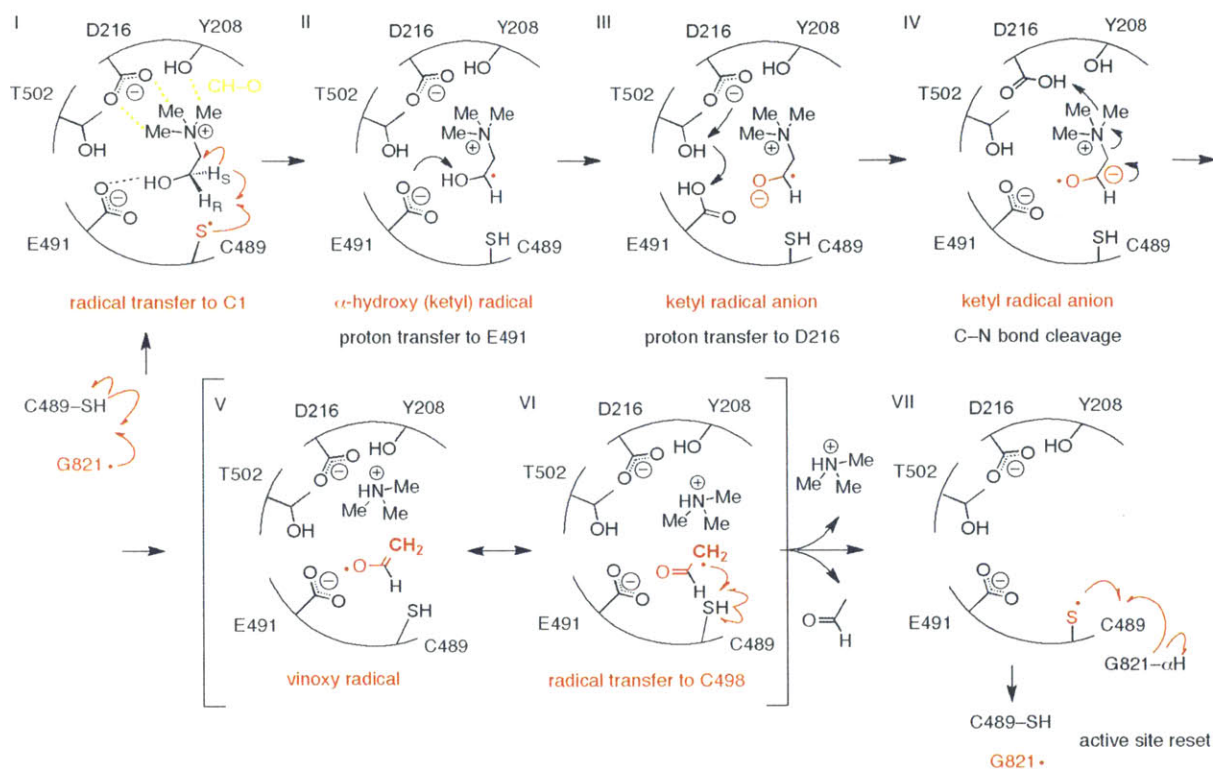


Figure IV.13. Structure-based mechanistic proposal for choline cleavage by CutC.

Choline binds to the active site residues through CH–O interactions (TMA moiety) and hydrogen bonds (C1 hydroxyl). The initial Gly to Cys results in a transient C498 thiyl radical within the active site. Choline cleavage begins by pro-S hydrogen atom abstraction from C1 (I). Deprotonation of the ketyl radical by E491 (II) leads to proton transfer to D216 mediated by T502 (III). Heterolytic C–N bond cleavage (IV) is facilitated by the resonance structure of the ketyl radical anion, which allows direct formation of the double bond to yield a vinyloxy radical (V), which is resonance stabilized by the acetaldehyde C2 radical. Hydrogen re-abstraction at C2 (VI) regenerates the thiyl radical. Release of TMA and acetaldehyde is permitted only after protonation and radical transfer are complete. The active site is now reset for a new round of catalysis (VII). Alternate mechanisms can be drawn with concerted or distributed proton transfer steps, hydrogen and proton transfer, or C–N bond cleavage and proton transfer.

A similar hydrogen-bonding network to the one described above is found in GDH. In GDH, this network is proposed to facilitate proton exchange between the Cys loop glutamate (analogous to E491) and a distal histidine (in place of D216) that is deprotonated by the eliminated hydroxide to yield water. It is proposed that after C–O bond cleavage, proton transfer between this histidine and glutamate could occur with a very low energetic barrier (0.7 kcal/mol) (20) over a distance of ~ 4.2 Å. In the choline-bound CutC structure, however, the proposed proton donor (E491) and acceptor (D216) are 5.8 Å apart, and a ~ 3 Å movement of the D216

side chain to reach E491 would be impeded by the presence of T502. Thus, we favor a role for T502 in this transfer based on the structure.

The GREs GDH and CutC are proposed to generate an initial C1 α -hydroxy radical similar to that proposed for the AdoCbl-dependent DDH and EAL; however, the mechanistic steps proposed for the AdoCbl enzymes are quite different. These enzymes have been studied extensively and a migration mechanism has been consistently supported by numerous computational methods (18, 19, 75). Accompanying biochemical experiments (76-78) have also supported migration in these systems and suggest why these enzymes might work hard to prevent an alternate, simpler mechanism such as elimination. AdoCbl enzymes are proposed to generate a transient 5'-dA \bullet , a much more oxidizing species than a thiyl radical (19), allowing 5'-dA \bullet to easily oxidize substrates, but a problem arises when one considers how the product-radical species can oxidize 5'-dA-H to reform 5'-dA \bullet . It has been proposed for ethanolamine ammonia-lyase that 1,2-migration facilitates re-abstraction of a hydrogen from 5'-dA-H as the product-based radical intermediate generated in this case is not stabilized by resonance and thus much more reactive (19). If base-catalyzed elimination were to occur in one of these enzymes, the vinoxy radical formed would be stabilized by resonance [C-H bond enthalpy, \sim 95.5 kcal/mol (79)] as proposed in the mechanism above and thus unable to effectively regenerate 5'-dA \bullet [5'-dA-H bond enthalpy, 99.9 kcal/mol (80)]. Such side reactions are observed with mechanism-based inhibitors, (81-83) all of which form low-energy radical species incapable of reoxidizing 5'-dA-H (84), and in mutants with natural substrates (85). This stalling of hydrogen atom reabstraction from 5'-dA-H effectively inactivates the enzyme, necessitates exchange of the Cbl cofactor and potentially causes radical damage within the cell.

Although we believe the active site of CutC is more likely to accommodate a base-catalyzed elimination like that proposed based on computational studies for GDH, our structures and biochemical experiments cannot rule out a migration mechanism as proposed for AdoCbl systems. Such a mechanism could make use of some of the very features we have ascribed to the radical mechanism, including E491 and D216 as a general base/acid pair, the putative proton relay network, and CH-O bonds for holding on to the trimethylammonium moiety. There are, however at least five reasons a base-catalyzed elimination makes sense in the context of a GRE system:

(1) CutC shares an overall protein architecture and a number of specific active site features with GDH. Although there has not been any detailed mechanistic characterization of this enzyme as of yet, an elimination mechanism is favored by DFT calculations. These calculations were based on the glycerol-bound structure of GDH and found reasonable transition state energies for C–O bond cleavage facilitated by general acid and base catalysis; no reasonable transition state for a migration mechanism was found. The key active site residues in this system are a glutamate on the Cys loop, which acts as a proton acceptor, and a histidine in the rear of the active site, which serves as a proton donor for the leaving hydroxide. CutC maintains the Cys loop architecture with virtually no changes. There are dramatic alterations in the rear of the active site that allow for binding of the TMA moiety; however, the proton donor/acceptor system seem to have been largely retained, with T502 serving to increase the distance over which the proton can be moved.

(2) In contrast to 5'-dA• in AdoCbl enzymes, the glycylyl and thiyl radicals in GREs are much more easily regenerated by product radicals, even a more stabilized protein radical species [the BDE of Cys S–H is ~87 kcal/mol (86) and BDE of the acetaldehyde radical is ~95.5 kcal/mol (79)], removing the necessity for the more complicated migration reaction.

(3) Choline appears to be ideally suited to the elimination of TMA from the choline C1-radical, as evidenced by the fact that the uncatalyzed elimination following oxidation by radiolysis or H₂O₂/Ti(III) proceeds very rapidly in solution, apparently due to the ease of deprotonation of the choline alcohol (87). The rate-limiting step in the CutC reaction is not currently known, but experiments with deuterated choline analogs are ongoing to address this question.

(4) In an elimination mechanism, once the choline C1 radical is formed, the enzyme would act primarily to facilitate movement of protons from C1 to TMA, rather than guiding migration of the TMA group. If migration of the TMA group were key to catalysis, one would expect residues that interact with the TMA group to be extremely important. However, we find that mutants that eliminate CH–O interactions with choline are not completely inactive despite the loss of the ability to hold on to the trimethylammonium moiety. The catalytic efficiency of these mutants is reduced substantially (>100 fold), so it is likely that coordination of the TMA moiety by CH–O bonds does assist the C–N cleavage, perhaps by orienting the TMA moiety for protonation. Future structural and biochemical studies will address this question.

(5) The bulky nature of the TMA moiety and its coordination by active site residues on the far side of the active site from the Cys loop also stands in contrast to the situation in AdoCbl enzymes where a small amino or hydroxyl moiety is thought to undergo migration through the action of hydrogen bond donors and acceptors all around the active site. The methyl groups of TMA are coordinated only by residues on the far side of the active site, and any movement toward C1 would disrupt these interactions and, furthermore, cause clashes of the TMA methyl groups with T502 and E491. Movement of C1 toward TMA would be equally disruptive as this would pull the C1 hydroxyl/ketyl radical away from the hydrogen bond donors on the Cys loop.

Radical enzymes have adopted many different strategies to bind substrates and perform chemistry not available by purely polar mechanisms. We propose that CutC uses electrostatic and CH–O interactions to recognize choline and has an alternate proton relay system in order to maintain the conserved chemical reaction seen in the related GRE eliminase GDH. Our studies with mutants have begun to pick apart this mechanism, but much work remains to be done to verify the roles we have suggested for active site residues. In particular, computational studies have the potential to guide mechanistic understanding, especially in systems such as CutC where there are numerous experimental difficulties. These techniques rely on accurate active site models to define the substrate geometry and active site residues in play. With structures of wild-type CutC and catalytically impaired mutants available, a number of computational and experimental techniques are available for future studies.

IV.E Materials and Methods

Protein purification.

Wild-type CutC with an 18 or 52 residue N-terminal truncation and an N-terminal His₆-tag and thrombin cleavage site were expressed and purified as previously described (17). The N-terminal truncations remove a putative microcompartment localizing sequence that has been previously noted to induce aggregation in ethanolamine ammonia-lyase (88). Full-length and truncated constructs have very similar activity, but the truncated proteins are more soluble and less prone to aggregation (17). The 52 residue truncated construct was favored for crystallization as it removes additional residues that are not conserved that appear to be unstructured. Site-directed mutagenesis to introduce the E491Q, Y208F, T502A and T502S mutations was

performed as previously described (17) with the CutC construct containing a 52 residue N-terminal truncation and a N-terminal His₆-tag and thrombin cleavage site. Mutant proteins were expressed and purified identically as wild type. Wild-type and mutant CutC was purified by anaerobic size-exclusion chromatography on a HiLoad 26/600 Superdex 200 column (GE Healthcare). This step separated the dimer from a minor population of aggregates and improved the activation results for several mutants. The final buffer for all proteins contained 50 mM potassium phosphate pH 8.0, 50 mM KCl, and 10% (v/v) glycerol. CutD was purified as previously described (17).

Crystallization and structure determination of wild-type CutC.

Initial screening of unactivated, wild-type CutC protein with the intact, N-terminal His-tag and 18 residue truncation was performed aerobically with the aid of an Art Robbins Phenix micro-pipetting robot and a Formulatrix Rock Imager. Numerous initial conditions were found, and optimization yielded well solution containing 17% (w/v) polyethylene glycol (PEG) 8000, 0.3 M lithium chloride, and 0.1 M Tris pH 8.0. Diffraction-quality crystals were obtained in hanging drop vapor diffusion trays at 21 °C. Protein at 15 mg/mL in buffer containing 50 mM potassium phosphate pH 8.0, 50 mM KCl, and 10% (v/v) glycerol was mixed with well solution in a 1:1 ratio. Plate-like crystals formed after 2-3 days of equilibration and grew to maximum size over one week. Crystals were cryoprotected by soaking for 15-30 min in solution containing 20% (v/v) glycerol, 25% (w/v) PEG 8000, 0.5 M lithium chloride, 0.1 M Tris pH 8.0. Choline was not added to the protein prior to crystallization, but initial analysis of the structure revealed density for choline, suggesting it is bound tightly. To ensure choline was bound at full occupancy in the wild-type protein, 10 mM choline was included in the cryoprotection solution. Crystals of wild-type CutC with an N-terminal 52 residue truncation grew in the same conditions in 1-2 days, but were long rods.

Crystals were initially characterized using an in-house CuK α X-ray source and indexed in space group P2₁ with cell edges $a = 80 \text{ \AA}$, $b = 234 \text{ \AA}$, $c = 106 \text{ \AA}$, $\beta = 110^\circ$. High-resolution data were collected at the Advanced Photon Source beamline 24ID-C on a Pilatus 6M detector (Dectris). Data were indexed, integrated, and scaled in HKL2000 (89) (statistics in Table IV.3). The structure of CutC was solved by molecular replacement in the Phenix implementation of Phaser (90). The structure of AdoCbl-independent glycerol dehydratase (PDB code 1R8W, 37%

identity) (21) was used to construct a search model after trimming of non-identical side chains in Phenix Ensembler (91). A solution with four molecules per asymmetric unit was found with an initial R factor of 0.477 at 2.8-Å resolution. Several rounds of initial refinement in phenix.refine (91) with tight NCS restraints were sufficient to reduce R_{free} values below 0.4, and manual building of side chains further reduced R_{free} to ~0.3. NCS restraints were removed after initial refinement and addition of water molecules. Positional and B factor refinement continued at the full resolution until the model was complete (statistics in Table IV.3). All ligands were fit into difference density and verified with simulated annealing composite omit maps. Ligand parameter files were generated in Phenix eLBOW. Water molecules were placed automatically after ligands were refined and were verified manually. The final model contains all native residues at the C terminus in all four molecules. No density is observed for the first 51 residues in the 18 residue truncated construct, but all other residues are present and well ordered in the model. Despite different crystal contacts for each molecule in the asymmetric unit, the structures of each monomer were essentially indistinguishable, with root-mean-square deviations of 0.11-0.14 Å. Structural figures were made in PyMOL v1.4.1 (Schrodinger).

(figure next page)

Table IV.3. Data collection and refinement statistics for the wild-type CutC with choline.

	wild-type CutC
	choline
Data collection and	
Resolution (Å)	50-1.90 (1.93-1.90) ²
Completeness (%)	99.3 (99.1)
$\langle I/\sigma \rangle$ ¹	12.5 (1.60)
R _{sym}	0.09 (0.99)
Unique reflections	282491
Redundancy	6.7 (5.9)
CC1/2	(0.68)
Model refinement	
Resolution (Å)	50-1.90
R _{work} , R _{free} ³	0.14 (0.18)
Protein atoms (chains)	25093 (4)
Choline molecules	4
Other ligand molecules ⁴	6
Water molecules	2471
RMSD ⁵ Bond length (Å)	0.011
RMSD Bond angle (°)	1.21
Rotamer outliers (%)	0.71
Ramachandran plot (%)	
Most favored	97.42
Additionally allowed	2.27
Disfavored	0.31
Average B factors (Å ²)	
Monomer A	40.9
Monomer B	33.3
Monomer C	31.6
Monomer D	29.9
Choline/TMA	27.7
Other ligands ⁴	42.8
Water	41.9

¹Data were trimmed based on CC1/2 value rather than the $\langle I/\sigma \rangle$ value. Inclusion of highest resolution data improved the quality of the electron density maps

²Values in parenthesis are for the highest resolution shell.

³3% of reflections were set aside prior to refinement for cross validation (8500-13300 reflections).

⁴Glycerol is bound at the protein surface in wild-type CutC structures.

⁵RMSD, root-mean-square deviation

Table IV.4. Data collection and refinement statistics for mutant CutC proteins

	CutC T502A- choline	CutC E491Q	CutC Y208F choline	CutC T502S choline
Data collection and				
Resolution (Å)	50-1.85 (1.88-1.85) ²	50-1.70 (1.73-1.70)	50-1.95 (1.98 -1.95)	50-1.85 (1.88-1.85)
Completeness (%)	99.7 (99.2)	100 (100)	99.4 (99.4)	99.4 (96.8)
$\langle I/\sigma \rangle$ ¹	9.41 (1.80)	10.0 (2.15)	9.49 (1.68)	11.6 (2.54)
R _{sym}	0.14 (0.84)	0.12 (0.88)	0.11 (0.80)	0.12 (0.57)
Unique reflections	347306	446331	295446	344874
Redundancy	5.0 (4.5)	6.2 (5.9)	3.7 (3.5)	6.2 (4.5)
CC1/2	(0.51)	(0.657)	(0.56)	(0.52)
Model refinement				
Resolution (Å)	50-1.85	50-1.70	50-1.95	50-1.85
R _{work} , R _{free} ³	0.158 (0.208)	0.165 (0.202)	0.176 (0.212)	0.194 (0.219)
Protein atoms (chains)	25441 (4)	25688	25438 (4)	25398 (4)
Choline molecules	4	NA	4	4
Other ligand molecules ⁴	17	21	18	15
Water molecules	4719	4178	4584	3280
RMSD ⁵ Bond length (Å)	0.003	0.007	0.004	0.004
RMSD Bond angle (°)	0.629	1.080	0.849	0.828
Rotamer outliers (%)	0.74	1.21	0.26	0.81
Ramachandran plot (%)				
Most favored	96.68	97.45	97.11	97.48
Additionally allowed	3.10	2.27	2.67	2.27
Disfavored	0.22	0.28	0.22	0.25
Average B factors (Å ²)				
Monomer A	18.6	15.9	18.8	13.2
Monomer B	18.1	15.8	18.9	13.2
Monomer C	20.3	17.8	21.7	14.9
Monomer D	19.7	17.8	21.7	14.9
Choline/Water	21.3/-	-23.9*	21.3/-	13.0/-
Other ligands ⁴	35.8	32.9	35.5	29.0
Water	27.1	31.0	31.1	21.9

¹Data were trimmed based on CC1/2 value rather than the $\langle I/\sigma \rangle$ value. Inclusion of highest resolution data improved the quality of the electron density maps

²Values in parenthesis are for the highest resolution shell.

³3% of reflections were set aside prior to refinement for cross validation (8500-13300 reflections).

⁴Malonate from the cryoprotectant is bound at the protein surface in mutant CutC structures.

⁵RMSD, root-mean-square deviation

Crystallization and structure determination of CutC mutants.

Initial attempts to crystallize CutC mutants in the same condition as wild-type CutC were unsuccessful. Therefore, new crystals of CutC mutants (E491Q, T502S, T502A, or Y208F) with an N-terminal 52 residue truncation were identified in screening trays using the same setup as for wild type. Optimized crystals were grown using hanging drop vapor diffusion at 21 °C (room temperature). CutC mutant protein at 8 mg/mL in buffer containing 50 mM potassium phosphate pH 8.0, 50 mM potassium chloride, and 10 mM choline was mixed in a 1:1 ratio with well solution containing 1.0-1.2 M sodium malonate pH 7.0-8.0. Crystals were rod-like and grew within 7 days. Crystals were cryoprotected by brief transfer into a solution containing 3.4 M sodium malonate pH 7.0 and 10 mM choline and cryocooled by plunging in liquid nitrogen.

Crystals of CutC mutants were indexed in space group $P4_2$ with cell edges approximately $a = b = 230 \text{ \AA}$, $c = 79 \text{ \AA}$. All data were collected and processed as for wild-type CutC (statistics in Table IV.4). Initial data processing suggested the presence of merohedral twinning in all datasets with high twin fractions and $P4_222$ as the space group. Despite this complication, the structure of each mutant was solved unambiguously by molecular replacement in space group $P4_2$ in the Phenix implementation of Phaser (90) using the wild-type monomer as a search model. Four molecules are present per asymmetric unit as observed before for the wild type. Positional and B factor refinement was conducted with phenix.refine (91) (statistics in Table IV.4). A pseudo two-fold axis perpendicular to the four-fold axis results in the appearance of higher symmetry and twinning. Twin refinement resulted in strong phase bias, little improvement in R_{free} , and a refined twin fraction of 0.5. Thus, the structures were refined without consideration of twinning. Initially, refinement was performed only on the T502A structure as the initial electron density maps were best for this dataset. Once the refinement converged, each mutant was refined separately (with R_{free} flags propagated to the other datasets). NCS and reference model restraints were used initially to reduce overfitting of the model but were removed after convergence with little increase of the R_{free} values. Each final model contains all native protein residues in each chain. No density is observed for the His-tag, but 7-10 residues of the thrombin cleavage tag present in this construct is clearly visible in all chains. Choline and active site waters were modeled into omit maps at the end of the refinement process.

Activity assays.

EPR spectroscopy, LC-MS/MS analyses and spectrophotometric coupled assays for CutC activity were performed as previously described (17) with the modification that all activation assays contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, and 200 μ M sodium dithionite. Glycyl radical formation was carried out in the same manner for all proteins prior to the kinetic assay. All activation assays contained the components above with 10 μ M dimeric CutC wild type or mutant (52-amino-acid-truncated construct), 40 μ M CutD in a volume of 50-200 μ L. The CutC enzymes were made anoxic by sparging with argon for 15-30 min. All components except for CutC and AdoMet were combined inside the anaerobic chamber in order to allow CutD to be reduced. CutC and AdoMet were added after 20 min. The reactions were incubated at room temperature for approximately 60 min before being diluted and immediately used in the kinetic assay.

All kinetic assays contained 200 μ M NADH, 25 mM Tris HCl buffer pH 8, 50 mM NaCl, activation mixture (as indicated below), YADH (as indicated below) and choline (as indicated below) in a total volume of 200 μ L. The assay was carried out in triplicate (except for T502A, D216N and Y208F/D216N which are single experiments) in a 96-well plate and the NADH absorbance at 340 nm was monitored for 5 min. The following conditions were used for each protein sample:

Wild type CutC: The activation assay was diluted 2000 fold (5 nM dimeric CutC), 0.3 μ M YADH, choline 0-10 mM.

Y208F: The activation assay was diluted 50 fold (0.2 μ M dimeric CutC), 2 μ M YADH, choline 0-10 mM.

Y506F: The activation assay was diluted 67 fold (0.15 μ M dimeric CutC), 0.4 μ M YADH, choline 0-20 mM.

Y208F&Y506F: The activation assay was diluted 5 fold (2 μ M dimeric CutC), 4 μ M YADH, choline 0-50 mM.

D216N: The activation assay was diluted 1.25 fold (8 μ M dimeric CutC), 6 μ M YADH, choline 2-150 mM, absorbance was measured over 14 min. The activity was very low, so the kinetic parameters are only estimates.

Y208F/D216N: The activation assay was diluted 1.25 fold (8 μ M dimeric CutC), 6 μ M YADH, choline 2-150 mM, absorbance was measured over 14 min. The activity was very low, so the kinetic parameters are only estimates.

T502A: The activation assay was diluted 2.5 fold (4 μ M dimeric CutC), 6 μ M YADH, choline 2-150 mM, absorbance was measured over 14 min. The activity was very low, so the kinetic parameters are only estimates.

T502S: The activation assay was diluted 500 fold (20 nM dimeric CutC), 0.4 μ M YADH, choline 0-20 mM

F395L: The activation assay was diluted 2 fold (5 μ M dimeric CutC), 4 μ M YADH, choline 0-10 mM

F395Y: The activation assay was diluted 67 fold (0.15 μ M dimeric CutC), 0.4 μ M YADH, choline 0-20 mM.

For E491A and E491Q mutant proteins, kinetic assays were impossible due to the very low turnover. Instead, a TMA detection assay was used to determine whether TMA was produced by these mutants. 20 mM choline was added to the activation mixture (including all components necessary for multiple rounds of activation) and incubated overnight (16 hours). 4 μ L of this reaction were diluted into buffer containing 20 % 5 mM ammonium formate, pH 4 and 80 % acetonitrile. This mixture was centrifuged at 13000 rpm for 10 min to remove particulates, then 3 μ L were analyzed by LC-MS/MS using the method for underivatized TMA detection (92).

IV.F Acknowledgements

M.A.F. is supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. 0645960. C.L.D. is a Howard Hughes Medical Institute (HHMI) Investigator. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by award GM103403 from the National Center for Research Resources at the National Institute of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. E.P.B. is supported by the Richard and Susan Smith Family Foundation Award for Excellence in Biomedical Research and the Packard Fellowship for Science and Engineering. S.B. is supported by an HHMI Graduate Fellowship.

IV.G References

1. Zeisel SH & da Costa KA (2009) Choline: an essential nutrient for public health. *Nutr Rev* 67(11):615-623.

2. Bain MA, Fornasini G, & Evans AM (2005) Trimethylamine: metabolic, pharmacokinetic and safety aspects. *Curr Drug Metab* 6(3):227-240.
3. Wang Z, *et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63.
4. Wang Z, *et al.* (2014) Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-*N*-oxide. *Eur Heart J* 35(14):904-910.
5. Tang WH, *et al.* (2013) Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 368(17):1575-1584.
6. Sato H (2010) Fecal trimethylamine, lactate and volatile fatty acid concentrations and their relationships with diarrhea in newborn dairy calves. *J Vet Med Sci* 72(12):1551-1555.
7. Neill AR, Grime DW, & Dawson RM (1978) Conversion of choline methyl groups through trimethylamine into methane in the rumen. *Biochem J* 170(3):529-535.
8. Kelly RH & Yancey PH (1999) High contents of trimethylamine oxide correlating with depth in deep-sea teleost fishes, skates, and decapod crustaceans. *Biol Bull* 196(1):18-25.
9. Seibel BA & Walsh PJ (2002) Trimethylamine oxide accumulation in marine animals: relationship to acylglycerol storage. *J Exp Biol* 205(Pt 3):297-306.
10. Ackermann D & Schutze H (1910) The formation of trimethylamine by *Bacterium prodigiosum*. *Zentralb Physiol* 24:210-211.
11. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.
12. Toraya T (2003) Radical catalysis in coenzyme B12-dependent isomerization (eliminating) reactions. *Chem Rev* 103(6):2095-2127.
13. Raynaud C, Sarcabal P, Meynial-Salles I, Croux C, & Soucaille P (2003) Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci U S A* 100(9):5010-5015.
14. Jorda J, Lopez D, Wheatley NM, & Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179-195.
15. Wei Y, *et al.* (2014) A chemically competent thiosulfuranyl radical on the *Escherichia coli* class III ribonucleotide reductase. *J Am Chem Soc* 136(25):9001-9013.

16. Wei Y, *et al.* (2014) The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proc Natl Acad Sci U S A* 111(36):E3756-3765.
17. Craciun S, Marks JA, & Balskus EP (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes. *ACS Chem Biol* 9(7):1408-1413.
18. Semialjac M & Schwarz H (2002) Computational exploration of rearrangements related to the vitamin B12-dependent ethanolamine ammonia lyase catalyzed transformation. *J Am Chem Soc* 124(30):8974-8983.
19. Wetmore SD, Smith DM, Bennett JT, & Radom L (2002) Understanding the mechanism of action of B12-dependent ethanolamine ammonia-lyase: synergistic interactions at play. *J Am Chem Soc* 124(47):14054-14065.
20. Feliks M & Ullmann GM (2012) Glycerol dehydration by the B12-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B* 116(24):7076-7087.
21. O'Brien JR, *et al.* (2004) Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.
22. Lehtio L, Grossmann JG, Kokona B, Fairman R, & Goldman A (2006) Crystal structure of a glycyl radical enzyme from *Archaeoglobus fulgidus*. *J Mol Biol* 357(1):221-235.
23. Blaser M (2006) Activation and regulation of the 4-hydroxyphenylacetate decarboxylase system from *Clostridium difficile*. Ph. D. Thesis (Philipps-Universität Marburg, Marburg, Germany).
24. Kerfeld CA, Heinhorst S, & Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64:391-408.
25. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem Rev* 111(4):2487-2506.
26. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA. How the enzyme uses the Cys-418 thiol radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.
27. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.

28. Becker A, *et al.* (1999) Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Biol* 6(10):969-975.
29. Martins BM, *et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *J Am Chem Soc* 133(37):14666-14674.
30. Funk MA, Judd ET, Marsh EN, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci U S A* 111(28):10161-10166.
31. Desiraju G (1998) Distinction between the weak hydrogen bond and the van der Waals interaction. *Chemical Communications* (8):891-892.
32. Adhikari U & Scheiner S (2013) Magnitude and mechanism of charge enhancement of CH..O hydrogen bonds. *J Phys Chem A* 117(40):10551-10562.
33. Horowitz S & Trievel RC (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J Biol Chem* 287(50):41576-41582.
34. Horowitz S, *et al.* (2013) Conservation and functional importance of carbon-oxygen hydrogen bonding in AdoMet-dependent methyltransferases. *J Am Chem Soc* 135(41):15536-15548.
35. Musah RA, *et al.* (1997) Variation in strength of an unconventional CH to O hydrogen bond in an engineered protein cavity. *Journal of the American Chemical Society* 119(38):9083-9084.
36. Dougherty DA & Stauffer DA (1990) Acetylcholine binding by a synthetic receptor: implications for biological recognition. *Science* 250(4987):1558-1560.
37. Luzhkov VB, Österberg F, Acharya P, Chattopadhyaya J, & Åqvist J (2002) Computational and NMR study of quaternary ammonium ion conformations in solution. *Physical Chemistry Chemical Physics* 4(19):4640-4647.
38. Selvaraj B, Pierik AJ, Bill E, & Martins BM (2013) 4-Hydroxyphenylacetate decarboxylase activating enzyme catalyses a classical S-adenosylmethionine reductive cleavage reaction. *J Biol Inorg Chem* 18(6):633-643.
39. Peng Y, Veneziano SE, Gillispie GD, & Broderick JB (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J Biol Chem* 285(35):27224-27231.

40. Bianchi V, *et al.* (1993) Flavodoxin is required for the activation of the anaerobic ribonucleotide reductase. *Biochem Biophys Res Commun* 197(2):792-797.
41. Mulliez E, Padovani D, Atta M, Alcouffe C, & Fontecave M (2001) Activation of class III ribonucleotide reductase by flavodoxin: a protein radical-driven electron transfer to the iron-sulfur center. *Biochemistry* 40(12):3730-3736.
42. Knappe J, *et al.* (1969) Pyruvate formate-lyase reaction in *Escherichia coli*. The enzymatic system converting an inactive form of the lyase into the catalytically active enzyme. *Eur J Biochem* 11(2):316-327.
43. Birch OM, Fuhrmann M, & Shaw NM (1995) Biotin synthase from *Escherichia coli*, an investigation of the low molecular weight and protein components required for activity in vitro. *J Biol Chem* 270(32):19158-19165.
44. Szu PH, Rusczycky MW, Choi SH, Yan F, & Liu HW (2009) Characterization and mechanistic studies of DesII: a radical *S*-adenosyl-L-methionine enzyme involved in the biosynthesis of TDP-D-desosamine. *J Am Chem Soc* 131(39):14030-14042.
45. Layer G, Verfurth K, Mahlitz E, & Jahn D (2002) Oxygen-independent coproporphyrinogen-III oxidase HemN from *Escherichia coli*. *J Biol Chem* 277(37):34136-34142.
46. Cicchillo RM, *et al.* (2004) Lipoyl synthase requires two equivalents of *S*-adenosyl-L-methionine to synthesize one equivalent of lipoic acid. *Biochemistry* 43(21):6378-6386.
47. Brazeau BJ, Gort SJ, Jessen HJ, Andrew AJ, & Liao HH (2006) Enzymatic activation of lysine 2,3-aminomutase from *Porphyromonas gingivalis*. *Appl Environ Microbiol* 72(9):6402-6404.
48. Grove TL, Lee KH, St Clair J, Krebs C, & Booker SJ (2008) In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry* 47(28):7523-7538.
49. Grove TL, *et al.* (2013) Further characterization of Cys-type and Ser-type anaerobic sulfatase maturing enzymes suggests a commonality in the mechanism of catalysis. *Biochemistry* 52(17):2874-2887.
50. Ollagnier S, *et al.* (1997) Activation of the anaerobic ribonucleotide reductase from *Escherichia coli*. The essential role of the iron-sulfur center for *S*-adenosylmethionine reduction. *J Biol Chem* 272(39):24216-24223.

51. Chandor-Proust A, *et al.* (2008) DNA repair and free radicals, new insights into the mechanism of spore photoproduct lyase revealed by single amino acid substitution. *J Biol Chem* 283(52):36361-36368.
52. Unkrig V, Neugebauer FA, & Knappe J (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analogue hypophosphite. *Eur J Biochem* 184(3):723-728.
53. Knappe J, Neugebauer FA, Blaschkowski HP, & Ganzler M (1984) Post-translational activation introduces a free radical into pyruvate formate-lyase. *Proc Natl Acad Sci USA* 81(5):1332-1335.
54. Leuthner B, *et al.* (1998) Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* 28(3):615-628.
55. Andersson J, Bodevin S, Westman M, Sahlin M, & Sjoberg BM (2001) Two active site asparagines are essential for the reaction mechanism of the class III anaerobic ribonucleotide reductase from bacteriophage T4. *J Biol Chem* 276(44):40457-40463.
56. Kessler D, Herth W, & Knappe J (1992) Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. *J Biol Chem* 267(25):18073-18079.
57. Yu L, Blaser M, Andrei PI, Pierik AJ, & Selmer T (2006) 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* 45(31):9584-9592.
58. Demick JM & Lanzilotta WN (2011) Radical SAM activation of the B12-independent glycerol dehydratase results in formation of 5'-deoxy-5'-(methylthio)adenosine and not 5'-deoxyadenosine. *Biochemistry* 50(4):440-442.
59. Perez C, Koshy C, Yildiz O, & Ziegler C (2012) Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. *Nature* 490(7418):126-130.
60. Cheng J, Goldstein R, Gershenson A, Stec B, & Roberts MF (2013) The cation- π box is a specific phosphatidylcholine membrane targeting motif. *J Biol Chem* 288(21):14863-14873.

61. Du Y, *et al.* (2011) Structures of the substrate-binding protein provide insights into the multiple compatible solute binding specificities of the *Bacillus subtilis* ABC transporter OpuC. *Biochem J* 436(2):283-289.
62. Mellroth P, *et al.* (2014) Structural and functional insights into peptidoglycan access for the lytic amidase LytA of *Streptococcus pneumoniae*. *MBio* 5(1):e01120-01113.
63. Iwase S, *et al.* (2011) ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* 18(7):769-776.
64. Govindasamy L, *et al.* (2004) Structural and mutational characterization of L-carnitine binding to human carnitine acetyltransferase. *J Struct Biol* 146(3):416-424.
65. Jogl G, Hsiao YS, & Tong L (2005) Crystal structure of mouse carnitine octanoyltransferase and molecular determinants of substrate selectivity. *J Biol Chem* 280(1):738-744.
66. Lenz R & Giese B (1997) Studies on the mechanism of ribonucleotide reductases. *Journal of the American Chemical Society* 119(12):2784-2794.
67. Hans M, *et al.* (2002) Adenosine triphosphate-induced electron transfer in 2-hydroxyglutaryl-CoA dehydratase from *Acidaminococcus fermentans*. *Biochemistry* 41(18):5873-5882.
68. Buckel W & Golding BT (1998) Radical species in the catalytic pathways of enzymes from anaerobes. *FEMS microbiology reviews* 22(5):523-541.
69. Dawson RMC (1959) Data for biochemical research. (Clarendon Press, Oxford).
70. von Sonntag C (2006) *Free-radical-induced DNA damage and its repair* (Springer).
71. Persson AL, *et al.* (1997) A new mechanism-based radical intermediate in a mutant R1 protein affecting the catalytically essential Glu441 in *Escherichia coli* ribonucleotide reductase. *J Biol Chem* 272(50):31533-31541.
72. Lawrence CC, *et al.* (1999) High-field EPR detection of a disulfide radical anion in the reduction of cytidine 5'-diphosphate by the E441Q R1 mutant of *Escherichia coli* ribonucleotide reductase. *Proc Natl Acad Sci U S A* 96(16):8979-8984.
73. Persson AL, Sahlin M, & Sjoberg BM (1998) Cysteinyl and substrate radical formation in active site mutant E441Q of *Escherichia coli* class I ribonucleotide reductase. *J Biol Chem* 273(47):31016-31020.

74. Zipse H, *et al.* (2009) Structure of the nucleotide radical formed during reaction of CDP/TTP with the E441Q-alpha2beta2 of *E. coli* ribonucleotide reductase. *J Am Chem Soc* 131(1):200-211.
75. Smith DM, Golding BT, & Radom L (2001) Understanding the mechanism of B(12)-dependent diol dehydratase: a synergistic retro-push--pull proposal. *J Am Chem Soc* 123(8):1664-1675.
76. Retey J, Umani-Ronchi A, Sebl J, & Arigoni D (1966) [On the mechanism of the propanediol dehydrase reaction]. *Experientia* 22(8):502-503.
77. Retey J, Umani-Ronchi A, & Arigoni D (1966) [On the stereochemistry of the propanediol dehydrase reaction]. *Experientia* 22(2):72-73.
78. Valinsky JE & Abeles RH (1975) The formation of unhydrated propionaldehyde by dioldehydrase. *Arch Biochem Biophys* 166(2):608-609.
79. da Silva G, Kim CH, & Bozzelli JW (2006) Thermodynamic properties (enthalpy, bond energy, entropy, and heat capacity) and internal rotor potentials of vinyl alcohol, methyl vinyl ether, and their corresponding radicals. *J Phys Chem A* 110(25):7925-7934.
80. Wetmore SD, Smith DM, Golding BT, & Radom L (2001) Interconversion of (S)-glutamate and (2S,3S)-3-methylaspartate: a distinctive B(12)-dependent carbon-skeleton rearrangement. *J Am Chem Soc* 123(33):7963-7972.
81. Abend A, Bandarian V, Reed GH, & Frey PA (2000) Identification of cis-ethanesemidione as the organic radical derived from glycolaldehyde in the suicide inactivation of dioldehydrase and of ethanolamine ammonia-lyase. *Biochemistry* 39(20):6250-6257.
82. Toraya T (2000) Radical catalysis of B12 enzymes: structure, mechanism, inactivation, and reactivation of diol and glycerol dehydratases. *Cell Mol Life Sci* 57(1):106-127.
83. Finlay TH, Valinsky J, Sato K, & Abeles RH (1972) Studies on the mechanism of action of coenzyme B 12 . The formation of 5'-deoxyadenosine and B 12(r) in the reaction of dioldehydrase with chloroacetaldehyde. *J Biol Chem* 247(13):4197-4207.
84. Sandala GM, Smith DM, Coote ML, Golding BT, & Radom L (2006) Insights into the hydrogen-abstraction reactions of diol dehydratase: relevance to the catalytic mechanism and suicide inactivation. *J Am Chem Soc* 128(10):3433-3444.

85. Sun L, Groover OA, Canfield JM, & Warncke K (2008) Critical role of arginine 160 of the EutB protein subunit for active site structure and radical catalysis in coenzyme B12-dependent ethanolamine ammonia-lyase. *Biochemistry* 47(20):5523-5535.
86. Rauk A, Yu D, & Armstrong D (1998) Oxidative damage to and by cysteine in proteins: An ab initio study of the radical structures, CH, SH, and CC bond dissociation energies, and transition structures for H abstraction by thiyl radicals. *Journal of the American Chemical Society* 120(34):8848-8855.
87. Foster T & West PR (1974) Photolysis of aqueous solutions of hydrogen peroxide containing β -ammonio alcohols. *Canadian Journal of Chemistry* 52(21):3589-3598.
88. Akita K, *et al.* (2010) Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of *Escherichia coli*. *J Biochem* 147(1):83-93.
89. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data. *Methods in Enzymology*, Vol 276.
90. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
91. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
92. Martinez-Del Campo A, *et al.* (2015) Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6(2).

Chapter V

The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant

This chapter has been published in a similar form: Wei, Y., Funk, M. A., Rosado, L. A., Baek, J., Drennan, C. L., & Stubbe, J. (2014). The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proceedings of the National Academy of Sciences*, 111(36), E3756-E3765.

This chapter was written and edited with Yifeng Wei, Dr. Leonardo A. Rosado, Prof. Catherine L. Drennan and Prof. JoAnne Stubbe. The biochemical experiments presented in the text were performed by Yifeng Wei. The bioinformatics analysis was performed by Yifeng Wei and Dr. Leonardo Rosado. The reagents and proteins used in this study were produced by Yifeng Wei and Jiyeon Baek.

V.A Summary

Ribonucleotide reductases (RNRs) catalyze the reduction of the four ribonucleoside di- or triphosphates (NDPs or NTPs) to the corresponding deoxynucleotides (dNTPs). Class III RNRs are strictly anaerobic glycy radical enzymes (GREs) that carry a stable glycy radical cofactor in their protein backbone with which they initiate radical based dehydration and reduction chemistry. These enzymes are structurally related to class I and II RNRs and are thought to perform some similar chemical steps, although these enzymes use different radical initiators. Class III RNRs are also homologous to other GREs like pyruvate formate-lyase, with which they share strategies for radical generation, control, and storage, but have different chemical functionality. The class III RNRs studied to date couple NTP reduction with the oxidation of formate to CO₂. Within this Chapter, we report the cloning and heterologous expression of the *Neisseria bacilliformis* class III RNR that does not use formate as a co-substrate. We show that instead it catalyzes nucleotide reduction using the ubiquitous thioredoxin/thioredoxin reductase/NADPH reducing system, in a manner we propose is similar to class I and II RNRs, where two active site cysteines serve as the reductant. We present a structural model based on a crystal structure of a related enzyme from *Thermotoga maritima*, showing the overall architecture of this subclass of RNRs and the position in the active site of three conserved cysteine residues that we have implicated in catalysis. Our structural and biochemical data allow us to posit the existence of at least three subclasses of anaerobic RNR that may use different catalytic steps to perform the common overall reaction. Analysis of the distribution of RNRs within bacteria and archaea suggest that the form of class III RNR used in a particular organism is connected with the metabolism of that organism. The discovery of this RNR subclass unites class III and class I and II RNRs in both performing nucleotide reduction chemistry with disulfides, although whether this is a case of convergent or divergent evolution remains to be seen.

V.B Introduction

The class III ribonucleotide reductases (RNRs) are glycy radical enzymes (GREs) present in many strict and facultative anaerobes that catalyze the conversion of nucleotides to deoxynucleotides (1, 2) with a mechanism involving complex free radical chemistry. RNRs are largely responsible for providing the balanced pool of deoxynucleotides required for DNA synthesis and repair (3). The class III RNRs that have been characterized thus far obtain the reducing equivalents required to make deoxynucleoside triphosphates (dNTPs) from the

oxidation of formate to CO₂ (4). Here we report a second subtype of class III RNR from *Neisseria bacilliformis*, which can obtain its reducing equivalents from the thioredoxin (TrxA)/thioredoxin reductase (TrxB)/ NADPH system.

RNRs provide the only pathway for de novo biosynthesis of dNTPs (5). They share a structurally homologous active site architecture in the catalytic subunit (often denoted α across classes) and a partially conserved, radical-based reduction mechanism. RNRs have been isolated and characterized from all kingdoms of life and many DNA viruses. Based on the characterization of a number of diverse enzymes thus far, RNR are divided into three classes (I, II, and III) according to the metallocofactor used to initiate the radical-dependent reduction chemistry (6). The class I RNRs use cofactors generated by the reaction of reduced metals (Fe, Mn, and Fe/Mn) and O₂ and are present only in aerobic organisms. The class II RNRs use adenosylcobalamin in an O₂-independent reaction and are present in both aerobes and anaerobes. The class III RNR uses an O₂-sensitive glycy radical (2) situated on a conserved loop, the Gly loop, in the α protein (NrdD). This radical is generated by a separate activating enzyme (AE or NrdG) through radical *S*-adenosylmethionine (AdoMet)-[4Fe-4S]¹⁺ chemistry (7, 8). All three classes ultimately generate a thiyl radical on cysteine within a conserved loop, the Cys loop, within the active site of α . Class III RNRs are only found in facultative and obligate anaerobes. A second distinction between the three classes has been the source of the reducing equivalents for nucleotide reduction. In the class I and II RNRs, they are provided by a redoxin (thioredoxin, glutaredoxin, or the RNR-specific, thioredoxin-like NrdH), which is rereduced by thioredoxin reductase and NADPH (9–11). In contrast, for bacteriophage T4 (12), its Gram-negative host *Escherichia coli* (1), and the Gram-positive *Lactococcus lactis* (13), the only class III RNRs characterized in detail to date, nucleotide reduction is coupled to the oxidation of formate to CO₂ (4).

Formate in *E. coli* and *L. lactis* is provided by carrying out the fermentation of sugars to acetate and formate in a pathway involving pyruvate-formate lyase (PFL) (14, 15). For *E. coli* growing in the absence of electron acceptors, formate induces the formate-hydrogenlyase pathway in which it is converted to the waste products H₂ and CO₂ by formate dehydrogenase (FDH) and hydrogenase (16). However, there are many proteins annotated as NrdD present in diverse bacteria and archaea (17, 18) which do not possess PFL or generate formate as an intermediate or end product in their primary metabolism (19), suggesting that an alternative

reducing system for class III RNRs might be involved. Although there is considerable variability in the presence or absence of formate-producing pathways, thioredoxin-like proteins used by the class I and II RNRs are ubiquitous. This observation prompted us to carry out a bioinformatics search for candidate NrdDs that use disulfide chemistry similar to the class I and II enzymes.

The generic mechanism of nucleotide reduction by all three classes of RNRs can be divided into two half reactions: the radical initiation process and the reduction process (20, 21). In all RNR classes, nucleotide reduction is initiated by generating a 3'-nucleotide radical (22–24) (ii in Figure V.1A) by a transient, conserved, top face thiyl radical (25) (i) on the Cys loop in the active site. This reaction likely involves general base catalysis by a conserved glutamate (class I and II), and perhaps formate (class III) (26), which facilitates loss of water to form a ketyl radical (iii). In the class I and II RNRs, reduction of the ketyl radical to a 3'-keto-deoxynucleotide intermediate is accompanied by the oxidation of the conserved cysteines (27) on the bottom face of the nucleotide to a disulfide anion radical (28) (iv in Figure V.1B), which serves as the reductant for the ketonucleotide, forming a 3'-deoxynucleotide radical and a disulfide (v). Product formation is accompanied by regeneration of the top face thiyl radical (vi). Rereduction of the active-site disulfide by TrxA occurs by disulfide exchange with a pair of conserved cysteines on the C-terminal tail of α (Figure V.1D) (27).

(figure next page)

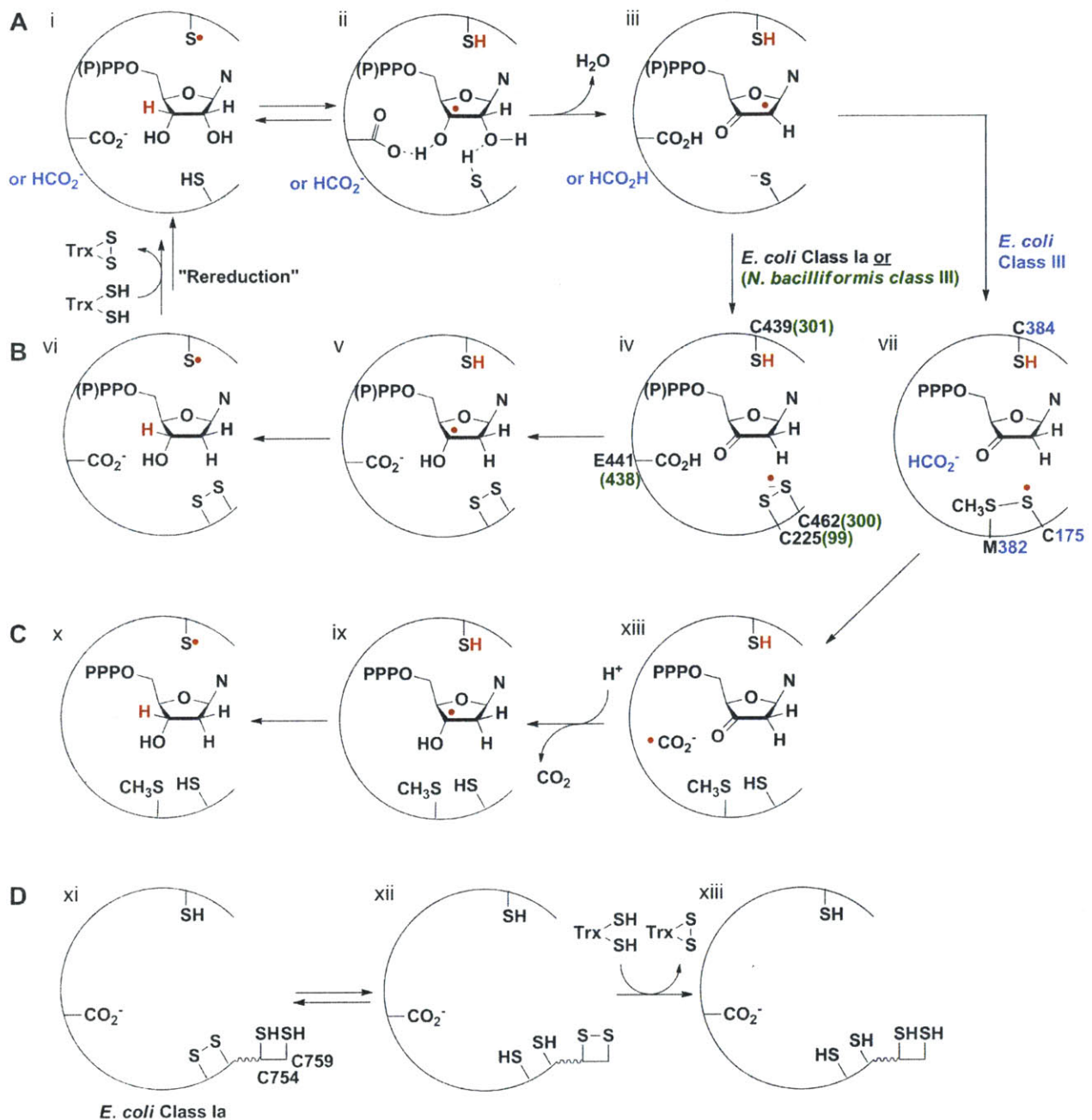


Figure V.1. Mechanistic model for nucleotide reduction by RNRs. (A) First half reaction common to all RNRs. **(B)** Second half reaction of *E. coli* class Ia and *N. bacilliformis* class III RNR. **(C)** Second half reaction of *E. coli* class III RNR (EcNrdD). M382 in EcNrdD RNR is located two residues from the top face thiyl radical (C384), in a position similar to the conserved N437 in *E. coli* class Ia RNR (EcNrdA), which makes a hydrogen bond to the 2'-OH group of the substrate. **(D)** Mechanistic model for re-reduction of the active site disulfide in class I and II RNRs via a pair of conserved cysteines on the C-terminal tail of α .

In NrdDs, only one of the disulfide-forming cysteines in the active site is conserved (21, 29, 30), and we recently showed (31) that reduction of the ketyl radical to the 3'-keto-deoxy-nucleotide is accompanied by the formation of a thiosulfuranyl radical (vii in Figure V.1B) between the bottom face cysteine thiyl radical and a methionine residue. The thiosulfuranyl radical, in equilibrium with the thiyl radical, then oxidizes formate to a carbon dioxide anion radical ($\bullet\text{CO}_2^-$) (21) (8), proposed to serve as the reductant for the 3'-keto-deoxy-nucleotide. In all classes, the 3'-keto-deoxynucleotide intermediate (iv in Figure V.1B or vii in Figure V.1C) is proposed to be reduced by proton-coupled electron transfer (32), with the source of the proton being the conserved glutamate in the class I and II RNRs, and unknown in the class III RNR.

Because of the role of this methionine residue in the reaction with formate in the *E. coli* NrdD (EcNrdD), we expected that it would be conserved in all formate-dependent NrdDs. However, just as the pathways for formate production are not conserved, our search in the RNRdb (33) showed that this methionine is missing in a set of NrdD sequences. In addition, all annotated NrdD sequences lacking this methionine residue invariably contain an additional cysteine residue immediately preceding the conserved thiyl radical on the Cys loop. This location may allow formation of a disulfide between the additional cysteine and the conserved bottom face thiol, thus allowing the reducing equivalents to be provided by chemistry similar to that in the class I and II RNRs.

To establish whether some class III RNRs use a formate-independent reduction strategy, a number of candidate class III RNRs were cloned and expressed. We now report the characterization of the NrdD and NrdG proteins from *N. bacilliformis* (NbNrdD and NbNrdG) (34, 35). This organism lacks the fermentative pathway terminating in PFL, a major source of formate for the class III RNRs studied to date, and its NrdD lacks the active site methionine. In addition, we were able to clone, express, isolate, and crystallize a related NrdD from the deep-branching thermophilic bacterium *T. maritima* (TmNrdD, 30% sequence identity with NbNrdD; Figure V.2). The mesophilic NbNrdD proved more amenable to biochemical studies, and we demonstrate that it is a GRE and show that, like the previously characterized class III RNRs, NTPs are substrates. We also show that formate is unable to provide the reducing equivalents to make dNTPs. This organism possesses a TrxA that has 61% sequence identity with *E. coli* TrxA, and activity can be reconstituted in vitro using the *E. coli* TrxA/TrxB/NADPH system. The X-ray crystal structure of TmNrdD reported here provides our framework for modeling the conserved

residues in this newly discovered class III RNR subtype and supports the hypothesis for the NbNrdD that three cysteines and a glutamate are located in the active site region where they can play a role in catalysis. The distribution and significance of this form of class III RNR are discussed.

Tm	MKVQYSFEREFEEELMSDLLSKYGYEMFQMDGLGDQLDVKFTEDFVRRGIIESTIDANAN	60
Nb	-----MIRLYPEQLNGKQLQFMHDIYSAHNAADG-----SKMDANAN	36
	:. *:* ..* :.* : . *	*.:*****
Tm	VRVTNISTYFIEISKPHTYLSLYRIWQMKMFMFGKGVAFDEFVEAQINGAVYLHDRHHAA	120
Nb	VTQKNIATMEAEIMKDFVQMNRAQVSRKIAEIEFDEATAREYIRQIEAHEIYVHD--ETS	94
	* .***: * * * . . : : * : * : * : * : * : * : * : * : * : * : * : *	:***: ..:
Tm	LMPYCFAYTLKPIVEKGLPFIKTIKSEPAKHLSTFIQHVIQFVMFASNQSSGAVGLPDDF	180
Nb	LKPYCVSVTLYPFLLDGLSKLGGE--SKAPKHLASFCGSFINLVFAISAQFAGAVATVEFL	153
	* ***. : * * * : .** . : * : .*** : * . * : * : * * :***. : * :	
Tm	VWMWYFVKKDLKEGLIPRDKLDWYIEQHFQILTYSLNQPIRTT--QSPYTNFTYLDRNYI	238
Nb	TYFDYFARKDYGEDYLETH--AAEVANHLQQVVYSINQFAARGYQSVFWNISVYDQYYF	211
	. : . * : * : * * . : . : : * : * : . * : * : * * : * : * : * : * : *	
Tm	KAIFEGERYPDGSLITDHVEDIIALQKHYWEWVSRERERQMFTFPVLTASLLYKDGKFLD	298
Nb	DAMFGDFVFPD--FSKPVWASVAKLQNFFLKWFNQERTKAVLTFPVVTAAMLTGGKCKD	269
	. * : * . : * * : . . : : * : * : * : * : * : * : * : * : * : * : *	
Tm	EDSARFINKINMKWQDTNWIYISDS IDAVASCCR LTSTQTLKKFSLSEEEEEKLGRMNS	358
Nb	GAFADQMAKELSEGNSFFVYLSDN PDSLASCCR LRN-----AIEDRTFS---YT	315
	* : * : : . * : * . * : : * * * * * . : * : * : . : :	
Tm	IGGSDLNIGSFKVIITVNLPRIALESGGDREKYLQILRHRVQLIKKALAAVREI IKERISE	418
Nb	LGAGGVATGSINVIITINMRLEQDGRD-----LAAEVAKIHRYQYAYRKLMEEYQAA	367
	: * : . : * : * : * : * : * : . . * . * * : : * * : * : * : *	
Tm	GLLPLYENGLMLLNRYQGTIGVTGVWESASIMGLTTEIDIGLKYTEEGEVFVDNVLDTIR	478
Nb	GMLPVYDAGFITLDKQFLTIGINGMAEAAESQGIKVG-----YNDYINQVQGRKTI F	421
	* : * : * : * : * : * : * : * : * : * . * : . * : : * : * . * : * *	
Tm	EEAEKGYHEYGFTFNIEQVPAEKAAVTLAQKDRFLFGEKQPFEIYSNQWVPLMANTDVLN	538
Nb	EANQAASKHYGVKFNTEFVPAENLGVKNKADGKADGYKVSRECYNSYFYVVEDEEINALD	481
	* : . : * : * . * * * * * : . * . * : * : . * : * : : * : *	
Tm	RIRYSG--KWDKKVSGGAILHINLGESFKTEEESFNMVKMIADMGVYFAFNTKISVCEG	597
Nb	KFLLHGKELVDWLDGGSALHLNLDEALPASG--YRSLLDIAAQTCNFCVNVTRITICNEC	540
	: : * : . : * : * : * : * : * : : . : : : * : * * * . * : * : * : *	
Tm	HAFYG---ERCPVCGKAKVDEY MRIVGYLVPVSAFN KERREIEYPRRQFYDSLTI RR--	651
Nb	GHIDKRTLHACSACGSHDIDYG TRVIGYLKRVSAFSS GRRKEHALRHYHREAQKQREAA	599
	: . * : * . * : * * : * : * * * * * * * : . * : . : : . *	

Figure V.2. Sequence alignment of *N. bacilliformis* (Nb) and *T. maritima* (Tm) NrdD. The alignment was created using Clustal Omega (91). Two important sequence features, the Gly loop and Cys loop are noted in bold.

V.C Results

Our bioinformatics analysis, leading to the identification of a previously unidentified NrdD subtype that couples nucleotide reduction to disulfide bond formation, is described below. To test our hypothesis, we attempted to clone, express, and purify six of the class III RNRs of this subtype, including the enzymes from *N. bacilliformis*, *T. maritima*, *Shewanella sediminis*, *Pyrococcus furiosus*, *Pseudomonas aeruginosa*, and *Schizosaccharomyces japonicas*. The NbNrdD and NbNrdG were soluble and could be obtained in reasonable amounts and thus became the focus of our attention.

NbNrdD Is a Glycyl radical Enzyme.

To generate active NbNrdD for biochemical studies, we incubated NbNrdD with NbNrdG and AdoMet in the presence of the acriflavin/bicine photoreduction system (31), resulting in the generation of a radical with an electron paramagnetic resonance (EPR) signal (Figure V.3A). The spectrum reveals a dominant hyperfine coupling constant of 40 MHz, proposed to be associated with the H α of glycine, consistent with its assignment as glycyl radical. Uniform labeling of NrdD with [²H]-glycine resulted in the collapse of the signal into a singlet (Figure V.3C), establishing the assignment. Additional spectral features (indicated with arrows in Figure V.3A), which persist even when the activation reaction is carried out in D₂O (Figure V.3B), are also visible for the PFL glycyl radical, although less well resolved (36). These features are attributed to hyperfine coupling with additional nonexchangeable protons, likely the α -protons of the two adjacent amino acids in the sequence (36). These interactions are predicted to be affected by the conformation of the peptide backbone, leading to variations in the glycyl radical EPR spectra between different GREs. The lack of exchange of the H α of glycine with D₂O, previously shown to occur with PFL (37), is similar to observations for the glycyl radical of EcNrdD RNR (38).

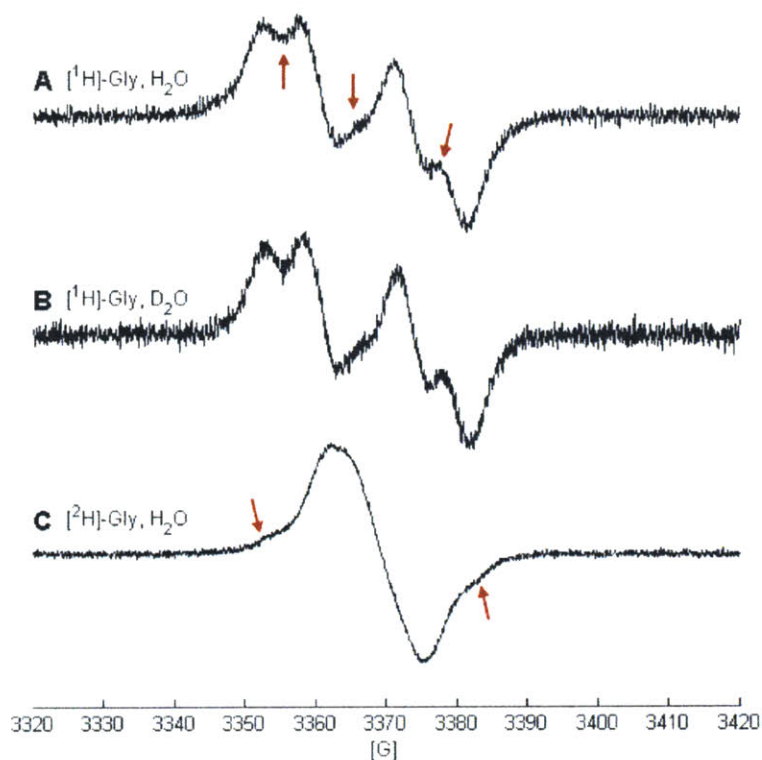


Figure V.5. X-band EPR spectra of the NbNrdD glycy radical. (A) NbNrdD in H₂O, red arrows indicate features arising from hyperfine coupling to nonexchangeable protons. **(B)** NbNrdD in D₂O. **(C)** [²H]-Gly-NbNrdD in H₂O. Red arrows indicate features possibly due to contaminating unlabeled NbNrdD.

NbNrdD Catalyzes CTP Reduction Using TrxA/TrxB/NADPH.

Our hypothesis is that reducing equivalents for nucleotide reduction by NbNrdD are delivered by a redoxin, similar to the class I and II RNRs (Figure V.1A and B). To identify a candidate redoxin for NbNrdD, we first carried out a BLAST search using the NbNrdD sequence. This search yielded a set of related sequences with ~50% pairwise identity in diverse organisms, including *S. sediminis* (γ -Proteobacteria), *Bacteroides ovatus* (Sphingobacteria), and *Clostridium citroniae* (Firmicutes). Examination of the redoxins present in these organisms revealed that only TrxA/TrxB is conserved. The high sequence identity between *N. bacilliformis* TrxA and *E. coli* TrxA (61% identity; Figure V.4), which has been used in assays for other class I and II RNRs (39–41), suggested that *E. coli* TrxA could be used in our activity assays.

```

Ec      MS-DKI IHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKI 59
Nb      MSSELI IHTTDVNFEQDVLNSDI PVLLDFWAPWCGPCKMIAPILDEVA AEYQGRLKIVKI 60
      ** : *** * .*: ***::* .*:***** *****:*** ***:*.:.*:
      .

Ec      NIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALS KGQLKEFLDANLA 109
Nb      NIDENEQTPAKFGVIRGIPTLMVFKDQNTATKVGALAKGQLTAFINASI- 109
      ***:* *..*:***::***:* :*****:****. *::*.:

```

Figure V.4. Sequence alignment of *E. coli* (Ec) and *N. bacilliformis* (Nb) TrxA. The alignment was created using Clustal Omega (91).

The assays were thus carried out with NbNrdD (0.25 glycyl radical per monomer) and *E. coli* TrxA/TrxB/NADPH, and the results are summarized in Table V.1. NbNrdD was active for reduction of CTP to dCTP with ATP as an effector, but nearly inactive for CDP reduction (~3% of the activity for CTP reduction; Table V.1). Catalytic activity was dependent on the presence of glycyl radical and TrxA (Table V.1). Formate failed to produce any dCTP. Unexpectedly, the activity was the same in the absence or presence of allosteric effectors (ATP or dATP; Table V.1). NbNrdD lacks the ATP cone domain that controls the activity of many RNRs by binding the activator (ATP) or the inactivator (dATP) (42). Thus, in NbNrdD, both ATP and dATP would be predicted to bind to the specificity site and activate nucleotide reduction. The activity that we have obtained with NbNrdD is ~0.24 s⁻¹ per glycyl radical, which is 20-fold lower than that of *E. coli* NrdD, which is ~4 s⁻¹ per glycyl radical (43). We hypothesize that the low activity and insensitivity to allosteric effectors is a result of *E. coli* TrxA functioning as a suboptimal reductant, making rereduction of the active site disulfide, rather than nucleotide reduction, rate limiting (44). Further study of the allosteric regulation of this enzyme will likely be facilitated by cloning, expressing, and using the *N. bacilliformis* Trx system in our assays.

Reaction conditions	Activity (U/mg)
Complete (CTP, ATP)	49
- AdoMet (no glycyl radical formed)	N.D.
-TrxA	N.D.
-TrxA, + formate (10 mM)	N.D.
-ATP	47
-ATP, + dATP (0.1 mM)	51
-CTP, +CDP (1 mM)	~1.5

Table V.1. Requirements for dCTP formation by NbNrdD (N.D. = not detected, < 3 turnovers per glycyl radical over 10 min).

The number of dCTPs formed per NADPH in the reaction mixture is 0.97 (Figure V.5), suggesting a 1:1 stoichiometry, in agreement with the proposal that the reducing equivalents are provided by the TrxA/TrxB/NADPH system (Figure V.1A and B). The ~26 turnovers per glycy radical that occur without addition of NADPH are attributed to the reduction of TrxA by residual DTT carried over from the NbNrdD storage buffer (~27 μM), which is essential for maintaining enzymatic activity.

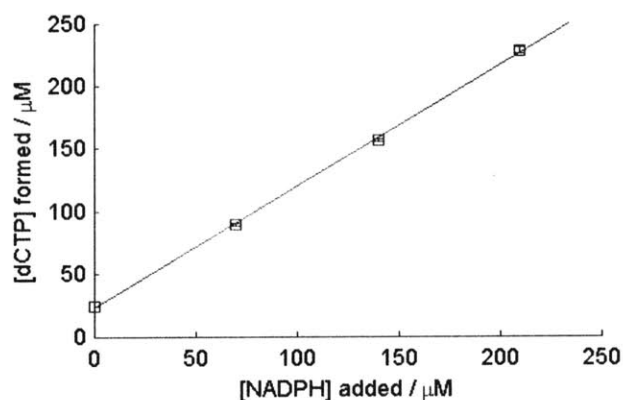


Figure V.3. dCTP production by NbNrdD. Amount of 5- ^3H -dCTP formed after incubation of NbNrdD with 5- ^3H -CTP, dATP, TrxA, TrxB, and limiting amounts of NADPH at 30 °C for 3 h. Stoichiometry of dCTP produced per NADPH added is 0.97. The concentration of NbNrdD glycy radical in the reaction is ~1 μM , and the ~26 turnovers per glycy radical that occur without addition of NADPH are attributed to the reduction of TrxA by residual DTT carried over from the NbNrdD storage buffer (~27 μM).

NbNrdD(C301A) Is Inactive, and Reaction of NbNrdD(C300A) with CTP Generates Cytosine.

To test our hypothesis that C301 forms the top face thiyl radical that initiates nucleotide reduction (Figure V.1A and B), the NbNrdD(C301A) mutant was generated. This mutant is inactive in dCTP and cytosine (Cyt) formation, despite having 0.25 glycy radical per dimer, consistent with our model. We propose that C300 in NbNrdD plays a role analogous to that of C462 in the *E. coli* class Ia α protein (EcNrdA) (Figure V.1B), donating reducing equivalents for nucleotide reduction by generating a disulfide with C99 (Figure V.1B). To test this hypothesis, we generated the NbNrdD(C300A) mutant and reacted it with CTP. The analogous C462A mutation in EcNrdA results in the generation of a 3'-keto-deoxycytidine intermediate (Figure V.6B) that decomposes to release Cyt (27).

The glycy radical of the NbNrdD(C300A) has an EPR spectrum identical to that of the WT-NbNrdD and is generated with a similar efficiency. Reaction of NbNrdD(C300A) with 5- ^3H -CTP leads to the time-dependent release of ~ 5.5 equivalents of 5- ^3H -Cyt per glycy radical (Figure V.6A), identified by HPLC (Figure V.7), with an initial rate of 9.3 U/mg ($\sim 2.5 \text{ min}^{-1}$ per glycy radical). No dCTP is detected, and the same amount of Cyt is produced in the presence or absence of the Trx system. A control with WT-NbNrdD shows no Cyt release either in the presence or absence of TrxA/TrxB/NADPH.

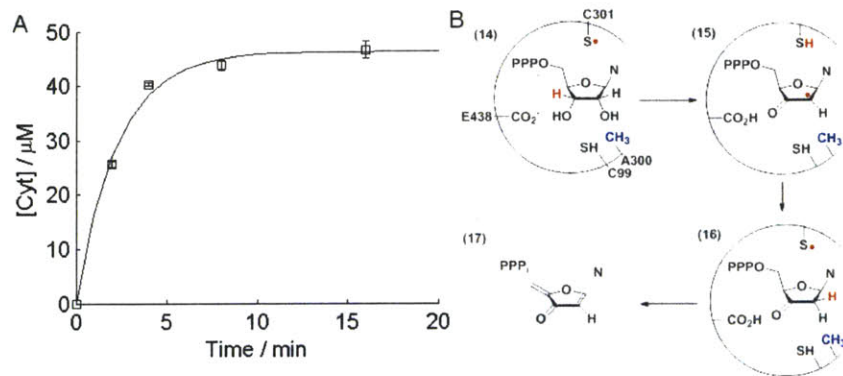


Figure V.6. (A) Time-dependent 5- ^3H -cytosine (Cyt) release in the reaction of NbNrdD (C300A) with 5- ^3H -CTP. The concentration of glycy radical in the reaction is $\sim 8 \mu\text{M}$. (B) Proposed mechanism for Cyt release by NbNrdD(C300A).

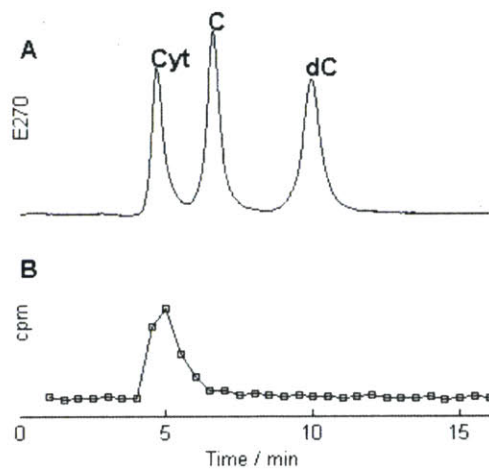


Figure V.7. HPLC detection of Cyt released in the reaction of NbNrdD(C300A) with 5- ^3H -CTP. (A) Standards monitored by absorbance at 270 nm. (B) 5- ^3H -Cyt detected by scintillation counting.

Crystal Structure of TmNrdD Allows Modeling of Active Site Residues in Redoxin-Dependent NrdDs.

Bioinformatics analysis (described subsequently) suggested that in addition to the three cysteines in the active site, a glutamate will also be present. To determine if these residues are located in the active site of this class of redoxin-dependent NrdDs, we wanted to obtain a structure of a representative of this NrdD subtype. Selenomethionine (SeMet)-labeled NrdD from *T. maritima* (TmNrdD), which is related to NbNrdD (30% sequence identity; Figure V.2), was successfully crystallized, and the structure was solved by single-wavelength anomalous dispersion to 1.64-Å resolution (Table V.2). We observe a (β/α)₁₀ barrel architecture similar to the T4 bacteriophage NrdD (29) [root mean square deviation (RMSD), 2.5 Å], including the C-terminal glycyl radical domain and Zn-binding site (Figure V.9A) but with four additional helices at the N terminus (Figure V.8A and Figure V.9B). The Cys loop containing essential cysteines C329 and C330 (equivalent to C300 and C301 in NbNrdD), however, is not present in its expected conformation within the barrel. In fact, residues 330–349 are not visible in the electron density map in either monomer, and SDS/PAGE analysis of the crystals indicates that protein cleavage occurs within this stretch of amino acids (Figure V.10). The molecular weights of the fragment bands (39 and 36 kDa) suggest that the missing density may be due to disorder around a single cleavage site rather than a missing stretch of residues. Regardless, cleavage within residues 330–349 has a dramatic affect; adjacent residues 320–329 and 350–365 move, and residues 350–365, which show no sequence or structural similarity to the Cys loop, now occupy the Cys loop position (Figure V.8B). Despite the absence of the Cys loop in our structure, the barrel architecture is intact, and the glycyl radical domain is ordered. The C terminus beyond the Gly loop is also ordered but adopts two distinct conformations in the two different molecules in the asymmetric unit (Figure V.8C).

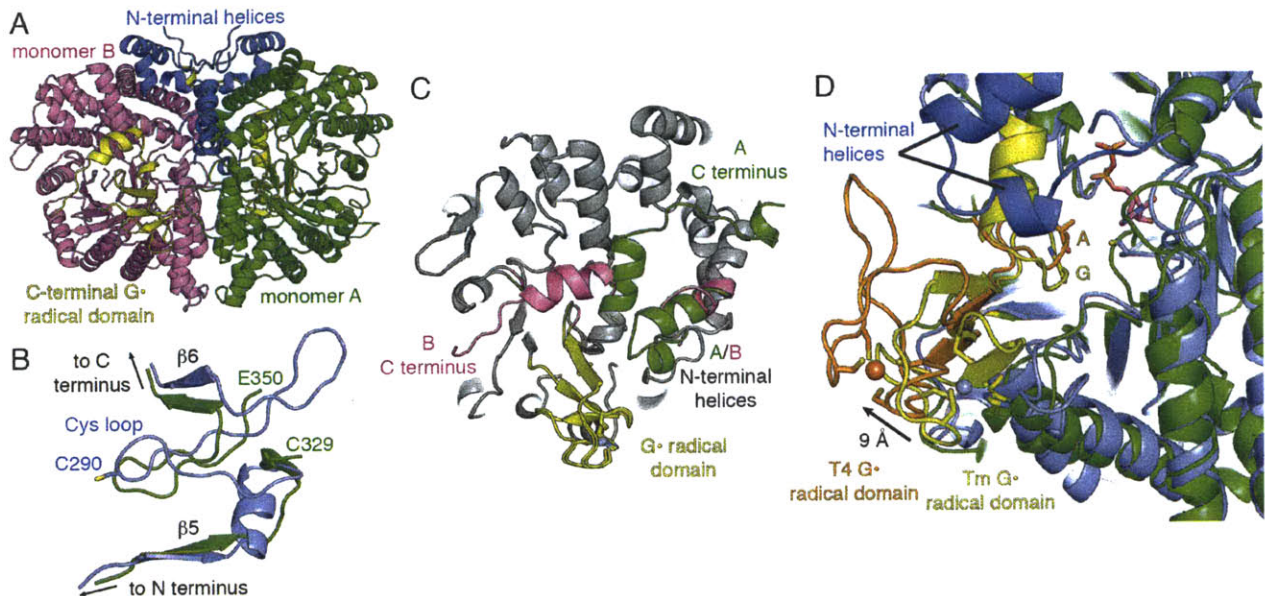


Figure V.8. Overall structure of *T. maritima* NrdD. (A) The structure of the dimer is shown with monomer A (green) and monomer B (magenta) joined by an interface that contains four N-terminal helices specific to TmNrdD (blue). (B) Superposition of residues 265–312 of the T4 bacteriophage structure (PDB ID code 1H79) (45), which includes the Cys loop (blue), with residues 314–329 and 350–369 of the TmNrdD structure (green). Residues 350–365 of TmNrdD occupy the position left vacant by the absence of the Cys loop (residues 324–334). (C) Monomers of TmNrdD differ by more than 1 Å in two regions: the tip of the N-terminal helices that contact the glycy radical domain and the C-terminal helix (green, monomer A; magenta, monomer B). The core of the protein is unchanged (gray). (D) Glycy radical domains of TmNrdD (yellow, Gly shown in sticks) in both chains are positioned further inside the barrel relative to that of T4 bacteriophage (orange, G580A in sticks). The C-terminal helix of TmNrdD is shown in its monomer A orientation (yellow). Zinc (TmNrdD, gray) and iron (T4 phage, orange) atoms, from Tm and T4 structures, respectively, are shown as spheres. CDP (magenta) is modeled based on the *T. maritima* class II RNR structure (PDB ID code 1XJN) (46). C and D are reproduced in Figure V.9 C-D as stereoisomers.

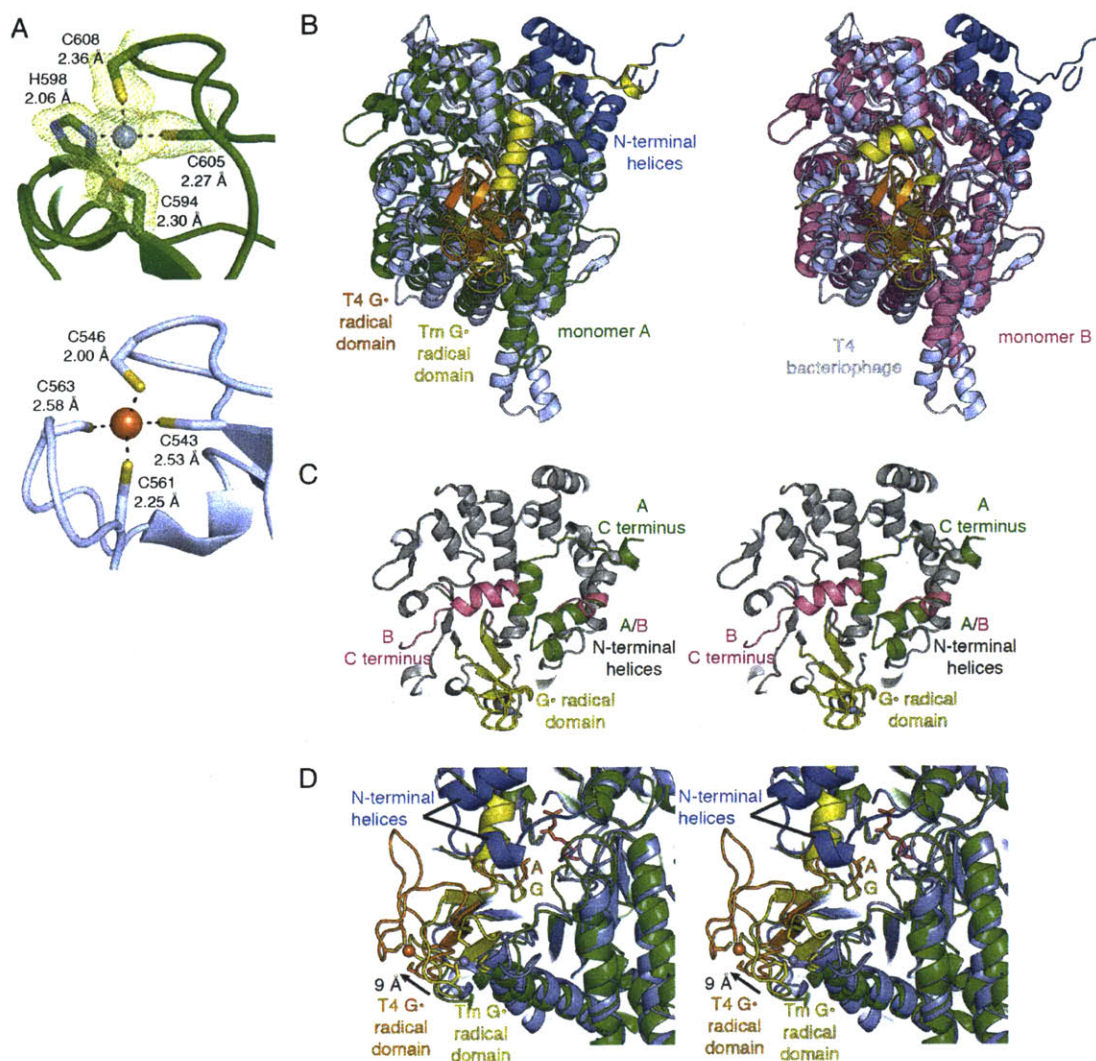


Figure V.9. Comparison of TmNrdD and T4 bacteriophage NrdDs. (A) The Zn-binding site of TmNrdD (green, monomer A) is similar to the T4 bacteriophage NrdD (light blue) (PDB ID code 1H79) (45) with the exception that H598 replaces C563. Composite omit maps around the Zn and ligating residues are contoured at 1σ (yellow). (B) Comparison of TmNrdD monomer A (green) and TmNrdD monomer B (magenta) to the T4 bacteriophage NrdD (light blue). The N-terminal helices in TmNrdD (dark blue) have no counterpart in the T4 NrdD. The T4 glycy radical domain (orange) and Tm glycy radical domain (yellow) are shifted as seen in Figure V.8D. (C) Stereoimage of Figure V.8C. Monomers of TmNrdD differ by more than 1 Å in two regions: the tip of the N-terminal helices that contact the glycy radical domain and the C-terminal helix (green, monomer A; magenta, monomer B). The core of the protein is unchanged (gray). (D) Stereoimage of Figure V.8D. Glycy radical domains of TmNrdD (yellow, Gly shown in sticks) in both chains are positioned further inside the barrel relative to that of T4 bacteriophage (orange, G580A in sticks). The C-terminal helix of TmNrdD is shown in its monomer A orientation (yellow). Zinc (TmNrdD, gray) and iron (T4 phage, orange) atoms, from Tm and T4 structures, respectively, are shown as spheres. CDP (magenta) is modeled based on the TmNrdJ structure (PDB ID code 1XJN) (46).

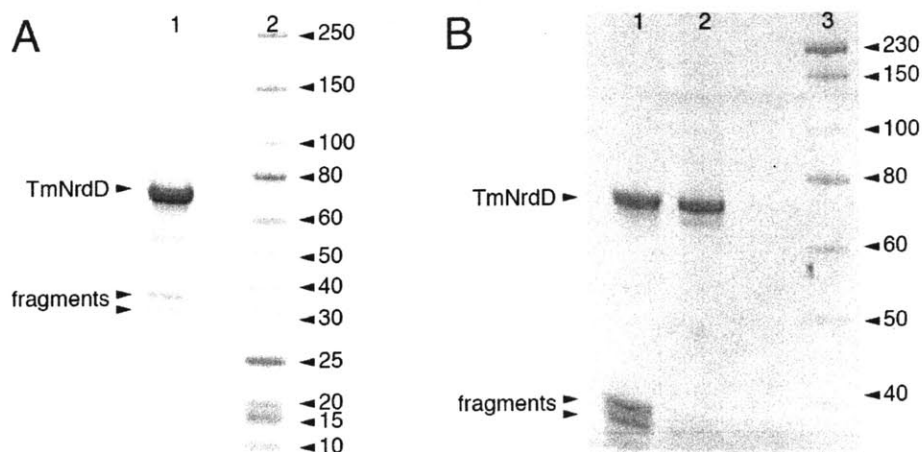


Figure V.10. TmNrdD fragments form during purification and are enriched by crystallization. (SDS/PAGE, 10% gels) **(A)** Purified TmNrdD contains intact enzyme (75 kDa) and a small amount of two fragments (~36 and ~39 kDa). Molecular weight markers in lane 2. **(B)** Unwashed crystals of TmNrdD (lane 1) show enrichment of the fragment bands, whereas the surrounding protein-precipitant solution (lane 2) is depleted in the fragmented protein. This differential suggests selective crystallization of the two fragments over intact enzyme. Molecular weight markers in lane 3.

Using the TmNrdD structure and T4 bacteriophage NrdD structure, we constructed a hybrid structural model (Figure V.11A). The Cys loop was modeled from T4 NrdD (T4 residues 287–293; equivalent to TmNrdD 327–333). Because there are no structures of a nucleoside triphosphate-bound RNR, CTP was modeled in the active site based on the position of CDP in the *T. maritima* class II RNR (TmNrdJ) (PDB ID code 1XJN) (46). Structural superpositions based on aligning residues in the β strands surrounding the active site yield all-atom RMSD values between TmNrdD and T4 NrdD of 0.9 Å and between TmNrdD and TmNrdJ of 2.0 Å.

In the TmNrdD model, C330 is positioned at the tip of the Cys loop to initiate catalysis (Figure V.11A). The modeled C329 is directly adjacent to C125 ($C\beta$ – $C\beta$ distance, 3.8 Å), sufficiently close to form the proposed disulfide (Figure V.1B), as occurs in EcNrdA ($C\beta$ – $C\beta$ distance, 4.1 Å) and TmNrdJ ($C\beta$ – $C\beta$ distance, 4.2 Å) (Figure V.11B). The precise orientation of the Cys loop and the placement of C329 in relation to the ribose of the CTP cannot be determined from this model, but the general locations of C329 and C125 are consistent with their proposed mechanistic role. In EcNrdA, reduction of the ketyl radical intermediate (Figure V.1A) is proposed to occur by hydrogen atom transfer from C225 from the bottom face of the nucleotide, followed by eventual disulfide formation with C462 (Figure V.1B), which is deeply buried on the innermost side of the active site. TmNrdD C125 is observed in the crystal structure to be

positioned equivalently to EcNrdA C225, consistent with it playing the same catalytic role. However, TmNrdD does not have a cysteine equivalent to EcNrdA C462 in the rear of the active site; S368 occupies this space in the structure. Instead, the location of C329 on the Cys loop requires formation of the disulfide at the front of the active site, in an orientation distinct from the class I/II enzymes, possibly in a position to facilitate its rereduction by redoxins (discussed below).

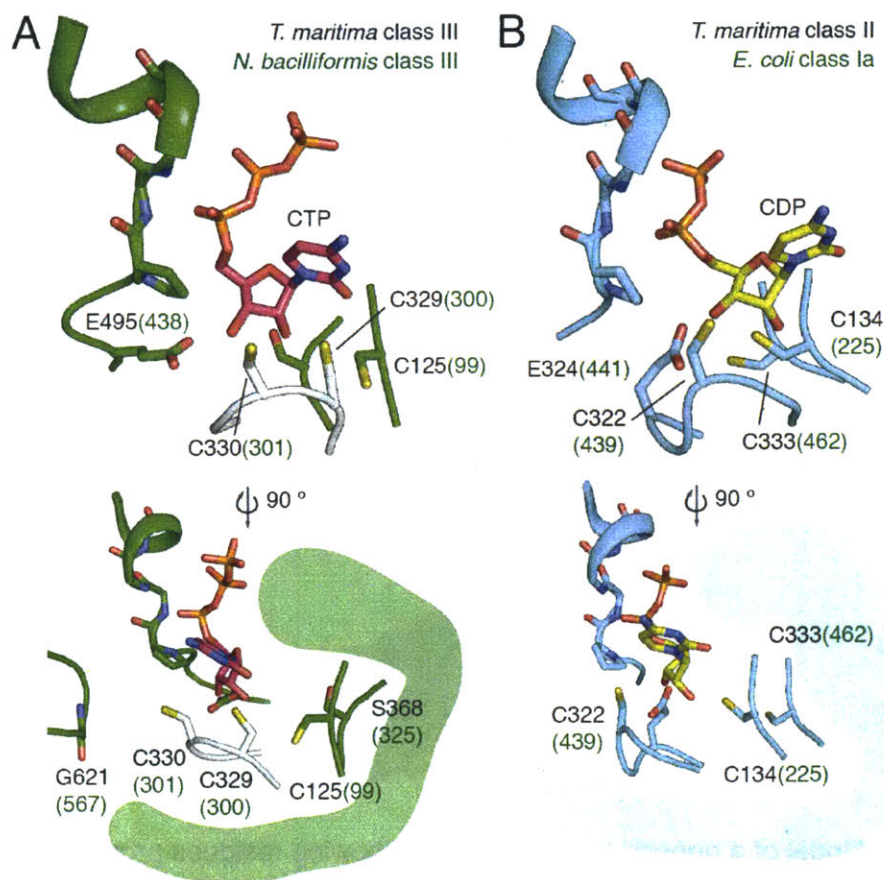


Figure V.11. Model for disulfide formation in NrdD. (A) The crystal structure of TmNrdD is shown in green (the inserted loop, residues 350–365, is not shown for clarity). The Cys loop based on bacteriophage T4 NrdD (PDB ID code 1H79) (45) has been modeled in white. CTP has been modeled based on TmNrdJ (magenta). The side view shows the position of the glyceryl radical loop. (B) The structure of CDP (yellow) bound TmNrdJ (PDB ID code 1XJN) (46).

The reduction of the 3'-keto-deoxynucleotide (Figure V.1B, 4) requires a proton in addition to the electron from the disulfide radical anion. We proposed from sequence alignments that residue E495 could function as the proton source, taking the place of E441 in EcNrdA. This glutamate is not conserved in the *E. coli*-type NrdD, but is conserved in the *T. maritima*-type

NrdD (Figure V.12 and Discussion). Here we find that the location of E495 in the TmNrdD structure overlaps quite well with the position of E441 (Figure V.11A). Unfortunately, the mutant of the corresponding residue in NbNrdD, E438Q, displayed low glycy radical content, and the enzyme was inactive in forming either dCTP or Cyt, which prevented us from carrying out experiments analogous to the *E. coli* NrdA(E441Q) mutant (28, 47).

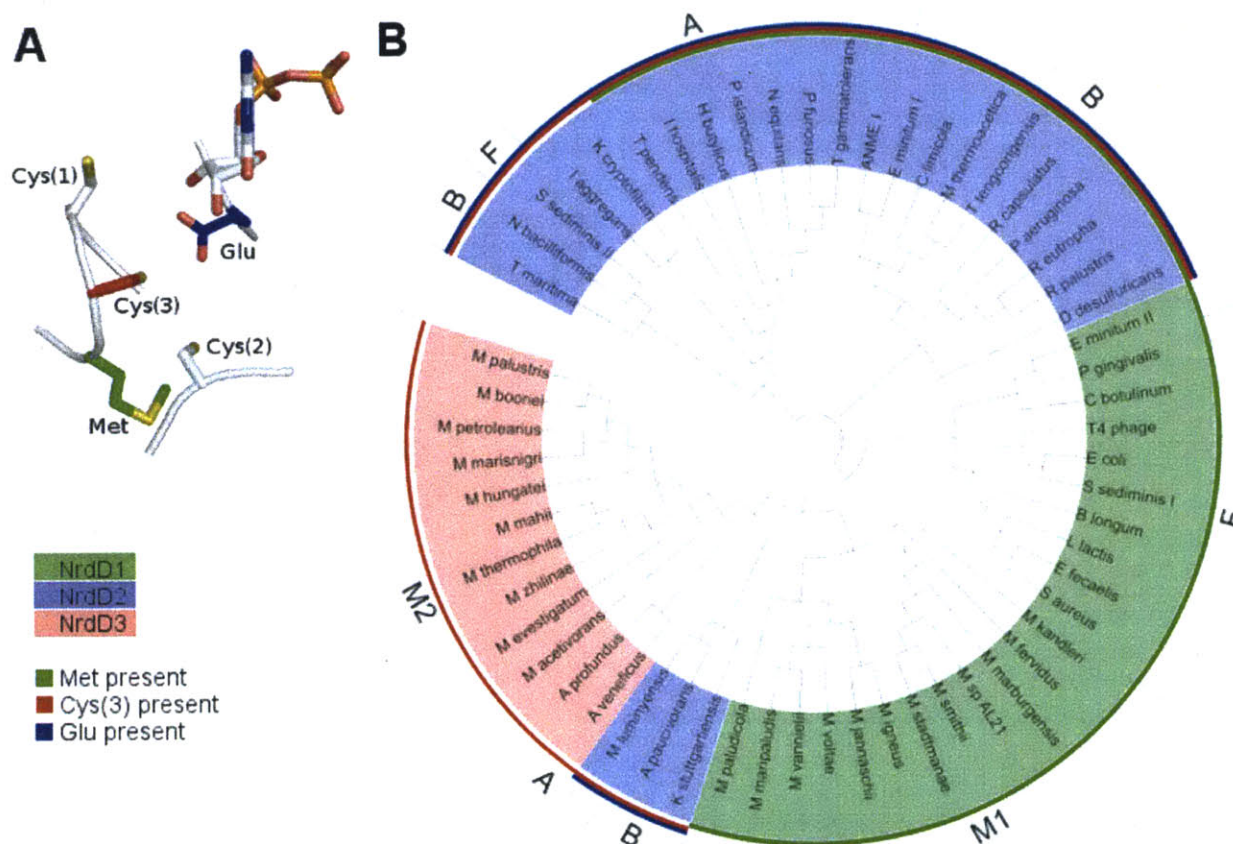


Figure V.12. (A) Model of a generic NrdD active site showing residues proposed to be important for chemistry. Cys(1) and Cys(2) are conserved in all NrdDs, and Met, Cys(3), and Glu are variously conserved in the different NrdD subtypes. **(B)** Phylogenetic tree of NrdDs. NrdD subtypes are indicated by highlighting (green = NrdD1, blue = NrdD2, red = NrdD3). Presence of Met, Cys(3), and Glu are indicated by colored bars (green, red, and blue, respectively) along the circumference of the plot. F, bacteria with PFL; B, other bacteria; M1, type I methanogens that use the cytosolic electron-bifurcating heterodisulfide reductase; M2, other methanogens; A, other archaea.

A major puzzle with respect to this newly discovered class III RNR subtype is how the active site disulfide, which must be formed on each round of catalysis, is rereduced by a redoxin. The class I and II RNRs require five cysteine residues for catalysis: three in the active site (Figure

V.1B) and two located in the C-terminal tail (Figure V.1D) that are involved in rereduction of the disulfide generated during dNDP (dNTP) formation so that multiple turnovers can occur. All NrdDs lack a C-terminal tail containing a pair of cysteines (27) that could function in this capacity. The TmNrdD structural model shows that C329 and C125 are found at the outer edge of the active site, in contrast to the deeply buried cysteine pair found in the class I/II enzymes. Nonetheless, a large conformational change would seem be required to allow the direct access of TrxA to the oxidized cysteine pair. The active site in our structure is buried primarily by the presence of the glycyl radical domain and two N-terminal helices specific to the TmNrdD-type enzymes. Although the active site appears buried, the $(\beta/\alpha)_{10}$ barrel architecture conserved in all classes of RNRs and all GREs is known to undergo conformational changes that can either close or expose the active site. In the class II RNR from *Lactobacillus leichmannii*, coenzyme B₁₂ binding causes a shift in a small number of residues exterior to the barrel, which closes the barrel and shields the cofactor from solvent (48). In PFL, the glycyl radical domain must exit the active site to become accessible to its partner AE for radical generation, and circular dichroism spectroscopy shows that binding of PFL to its AE is accompanied by enhanced enzyme conformational flexibility (49, 50). The recent structure of the GRE benzylsuccinate synthase (51) reveals a clamshell-like opening of the barrel, allowing release of the glycyl radical domain from the interior of the protein, which would again allow for glycyl radical formation. In this work, support for the proposal that the glycyl radical domain is flexible comes from the observation that the domain position in TmNrdD is shifted by 9 Å (~15° rotation) relative to that of T4 bacteriophage (Figure V.8D). Although the relevance, if any, of this shift is not established at this time, it illustrates the wide range of motion available to the glycyl radical domain in NrdD. Also notable for understanding active site access in TmNrdD is the observation that the glycyl radical domain C-terminal helix has two distinct conformations (Figure V.8C): in monomer A, the helix is extended and completely blocks active site access; in monomer B, the helix kinks and wraps around the protein, similar to other GREs, revealing an opening to the active site. Although this active site cavity is not large enough to accommodate TrxA, the flexibility observed in the C- and N-terminal helices, as well as in the glycyl radical domain, indicates the likelihood that further opening of the active site would permit rereduction by TrxA.

Bioinformatics Analysis Suggests Three Chemically Distinct NrdD Subtypes.

Further support for the existence of distinct subtypes of NrdDs was obtained by a bioinformatics analysis. A phylogenetic tree of NrdDs (Figure V.12B) was constructed using 59 sequences from the RNRdb (33), chosen to include phylogenetically and metabolically diverse bacteria and archaea. Analysis of the conserved residues proposed to be important for chemistry [the three active site cysteines Cys(1–3), Met, and Glu in Figure V.12A] led us to propose that there are three chemically distinct NrdD subtypes that we label NrdD1, NrdD2, and NrdD3. The top face Cys(1) on the Cys loop and the bottom face Cys(2) are conserved in all NrdDs, and Cys(3), Met, and Glu are variously conserved in the different NrdD subtypes. The phylogeny of these NrdDs suggests that horizontal gene transfer plays a large role in accounting for their distribution among the diverse bacteria and archaea, as postulated for all classes of RNR (18).

NrdD1 includes the previously studied bacteriophage T4, *E. coli*, and *L. lactis* enzymes that use formate as a reductant (4). Met, thought to play a role in the reaction with formate (Figure V.1 A and C) (31), is conserved in all NrdD1s. None of the NrdD1 sequences examined have Cys(3) and Glu. NrdD1s are present almost exclusively in fermentative bacteria, which have a PFL, and in hydrogenotrophic methanogens, where the F₄₂₀-dependent formate dehydrogenase provides a pathway for formate generation (52). In certain fermentative bacteria, like *E. coli* and *L. lactis*, PFL plays a role in energy metabolism by converting pyruvate to acetyl-CoA, which leads to ATP production through acetyl phosphate, and formate, which is a waste product (16). In some bacteria, PFL also plays a role in providing C1 intermediates for biosynthesis (53, 54).

NrdD2 includes the enzymes from *N. bacilliformis* and *T. maritima*, which are proposed to obtain reducing equivalents from a redoxin. In these proteins, three Cys(1–3) and a Glu, which enable nucleotide reduction by disulfide chemistry (Figure V.1A and B), are conserved. In these enzymes, Met is often present but is not conserved.

In NrdD2s lacking Met, this position is most commonly occupied by Ser, as in NbNrdD and TmNrdD, but also Ala, Glu, or Phe in some deeply rooted bacteria and archaea. The residue corresponding to the catalytic E441 in *E. coli* NrdA (Figure V.1A and B), which is expected to be close to the formate binding site, is replaced with either Ser or Thr in NrdD1s. However, in NrdD2s, it is replaced with Ile, Leu, or Phe, which may hinder the access of formate. In contrast to the restricted distribution of NrdD1, NrdD2 is present in non-methanogenic archaea and in bacteria with diverse types of anaerobic metabolism. In support of the nature of the proposed

reductant, several organisms have NrdD2s with associated redoxins, including *Acetobacterium woodii* (55), in the form of a C-terminal fusion to NrdD2, and *Desulfarculus baarsii* (56) and *Methanomassiliicoccus luminyensis* (57), which have a thioredoxin-like protein in the operon.

NrdD3, present only in certain methanogens and the closely related Archaeoglobi (58), shows important differences with respect to NrdD1 and NrdD2. Although they share up to 50% pairwise sequence identity with NrdD1s from other methanogens, they lack Met. Also, although they contain all three cysteines like the NrdD2s, they lack a glutamate. Additional observations support the designation of a third NrdD subtype and that these proteins will possess RNR activity. NrdD3s all contain the Gly loop consensus sequence, and some of them (e.g., *Archaeoglobus veneficus*) contain an N-terminal ATP cone domain. In some sequenced organisms, NrdD3 is the only annotated RNR (e.g., *Archaeoglobus profundus*, *Methanospirillum hungatei*), although other organisms also contain a NrdJ. Interestingly, some NrdD3s (*Archaeoglobus veneficus*, *Methanosarcina barkerii*) are found in the same operon as a thioredoxin-like protein.

The designation of the three NrdD subtypes largely correspond to the phylogeny of the protein (Figure V.12B), with the exception of a deeply rooted branch containing *M. luminyensis*, *Aminomonas paucivorans*, and *Kuenenia stuttgartiensis* NrdD2. An additional observation suggesting the existence of NrdDs with distinct types of chemistry is that, although it is uncommon for a single organism to contain two copies of the same NrdD subtype, there are many organisms that contain both NrdD1 and NrdD2 (e.g., *S. sediminis*, *B. ovatus*, and *E. minutum*). We hypothesize that these class III RNR variants are used under different anaerobic growth conditions, similar to the case of the class Ia and Ib RNRs in *E. coli*.

V.D Discussion

Because of their essential role in the de novo production of deoxynucleotides, RNRs are ubiquitous enzymes in nearly all cellular organisms and many viruses (18, 33, 59). The complex chemistry involved in nucleotide reduction requires initial generation of a transient thiyl radical. The class of RNR used by an organism reflects the mechanism of this radical formation and a combination of factors including the presence or absence of oxygen (60) and availability of metals (61). Here we report the discovery of a subtype of class III RNR in *N. bacilliformis* where, like the class I and II RNRs, nucleotide reduction is facilitated by a redoxin, which is a ubiquitous protein found in all organisms. This proposed reliance on a redoxin is unlike the class

III RNRs studied to date that couple nucleotide reduction to the oxidation of formate, a metabolite produced by some but not all organisms as part of their primary metabolism.

Biochemistry, Structure, and Bioinformatics Support the Existence of a Second NrdD Subtype.

The cloning of *N. bacilliformis* class III RNR was motivated by a bioinformatics analysis, which led to the identification of a potential second NrdD subtype containing three cysteines necessary to carry out nucleotide reduction through a mechanism analogous to that of the class I and II RNRs (Figure V.1A and B). Our biochemical investigations showed that NbNrdD catalyzes the reduction of NTPs, which are also the substrates of formate-dependent class III RNRs. dNTP formation required the presence of the TrxA/TrxB/NADPH system but not formate, and the 1:1 stoichiometry of NADPH consumption and dCTP production demonstrates that the reducing equivalents for dCTP generation are provided by the Trx system.

As predicted, the NbNrdD(C301A) mutant is inactive because no initiating thiyl radical can be produced. In support of the active site disulfide between C99 and C300 (Figure V.1B), our structural modeling using the related TmNrdD enzyme as the molecular scaffold is consistent with these cysteines being close enough to form a disulfide bond. Furthermore, the NbNrdD (C300A) mutant behaves like the EcNrdA RNR C462A mutant in that reaction with substrate (CTP in the case of NbNrdD) results in the formation of Cyt (Figure V.6B), showing that the mutant is competent for formation of a 3'-keto-dCTP intermediate but unable to carry out nucleotide reduction. Thus, our assays establish the presence of a second NrdD subtype that is distinct from *E. coli* NrdD in being able to use TrxA rather than formate as a reductant for nucleotide reduction.

Distribution of NrdD Subtypes Correlates with Metabolism.

Our bioinformatics study revealed that there are in fact three NrdD subtypes (NrdD1, NrdD2, and NrdD3), the distribution of which shows a striking correlation with the organism's anaerobic metabolism. Among bacteria, NrdD1 is localized almost exclusively in fermentative bacteria that use PFL. An interesting exception is the deep branching *Elusimicrobium minutum* (termite group 1; Figure V.12B) (62), which has two NrdD subtypes (NrdD1 and NrdD2) but no PFL: its NrdD1 operon contains a 10-formyl-tetrahydrofolate synthetase as a possible source of formate. In contrast, NrdD2 is present in bacteria carrying out a diverse range of anaerobic metabolism.

Although it is tempting to explain the distribution of NrdD subtypes by the availability of intracellular formate, a counter-example is provided by acetogens: although formate is an intermediate of acetogenesis (63), acetogens invariably contain NrdD2. Instead we hypothesize that the energy metabolism of the organism and the redox window that it inhabits determines whether it is more thermodynamically efficient to use formate or other reductants for nucleotide reduction. In fermentative bacteria, formate is produced by PFL as a waste product; thus, NrdD1 may be preferred. In contrast, in bacteria that carry out anaerobic respiration, formate is not known to be produced as part of their energy metabolism, and oxidation of any available formate from the environment can be coupled to the generation of ATP, which rationalizes why NrdD2 may be preferred. In acetogens, coupling of ribonucleotide reduction by NrdD2 to the TrxA/TrxB system could provide oxidized pyridine nucleotides that are substrates for energy conservation (55).

All archaea examined contain exclusively NrdD2, except for methanogens and their relatives, which contain either NrdD1 or NrdD3. In the case of the methanogens, the NrdD subtype used correlates with its mechanism for energy conservation (64), which is in turn related to its formate metabolism. NrdD1 is present only in type I hydrogenotrophic methanogens that carry out energy conservation by means of the cytosolic electron-bifurcating heterodisulfide reductase (64). The type I methanogens examined contain the F₄₂₀-dependent formate dehydrogenase (52) as a possible means of synthesizing formate, and many can carry out methanogenesis with formate as a substrate. Apart from ribonucleotide reduction, the other known role of formate in the primary metabolism of type I methanogens is in purine biosynthesis, in an ATP-dependent reaction with formate to form formylphosphate catalyzed by PurT (glycineamide ribonucleotide synthetase) (65, 66), suggesting the presence of intracellular formate. A case in point is *Methanocella paludicola* (rice cluster I) (67), a methanogen with cytochromes that nevertheless uses the type I pathway for methanogenesis (64), and contains NrdD1.

Other methanogens, including methylotrophic and acetoclastic methanogens that carry out energy conservation by other means, and their relatives, the sulfate-reducing Archaeoglobi, use NrdD3 instead. Unlike type I methanogens, these organisms are unable to use formate as a substrate for methanogenesis, and several sequenced members (e.g., *M. acetivorans* and *M. mazei*) do not contain formate dehydrogenase (68). They carry out purine biosynthesis in a

manner dependent not on formate, but on 10-formyl-tetrahydrofolate, catalyzed by PurN (glycineamide ribonucleotide synthase) (65), suggesting that formate may not be present in the cell. Although many NrdD3 operons contain a redoxin, methanogens lack a conserved thioredoxin reductase, and the origin of the reducing equivalents for NrdD3 is unclear, but could be linked to the heterodisulfide reductase system common to these organisms.

Several interesting methanogens and relatives also possess NrdD2. The obligate methylotroph *M. burtonii* (69) has both NrdD2 and NrdD3, suggesting that the source of reducing equivalents for these two RNRs may be different. The methanogen relative ANME-1 (70, 71), which carries out anaerobic methane oxidation, has NrdD2 but no NrdD3, in conjunction with its unique metabolism. *M. luminyensis* (57), which carries out methanogenesis from methanol but is a relative of the nonmethanogenic Thermoplasmatales, contains a deeply rooted NrdD2 with an associated redoxin in the operon. Because of the central role of formate and thiols in many anaerobic processes, the distribution of class III RNRs among different organisms may shed light on aspects of anaerobic biochemistry, particularly in organisms that are unculturable.

A Clue Regarding the Ancestral NrdD.

RNR has been proposed to provide the link between the RNA and DNA worlds (72, 73), with the class III RNR proposed to precede the class I and II enzymes. Given the central role of formate and thiols in many anaerobic processes, the identity of the original NrdD has implications regarding the metabolism of the first organism with a DNA genome. Unfortunately, the NrdD1 and NrdD2 sequences are too divergent to allow us to convincingly root the NrdD phylogenetic tree. However, we note that the residue Met (Figure V.12A) proposed to be involved in formate chemistry in NrdD1 (31) is also present in many but not all NrdD2s, possibly as an evolutionary relic, suggesting that NrdD1 may precede NrdD2. If true, this suggests that formate is present in the ancestral organism, possibly as an intermediate in an ancient Wood–Ljungdahl pathway (55).

V.E Materials and Methods

Materials and General Methods.

All chemical reagents were purchased from Sigma-Aldrich, unless otherwise indicated. Primers were purchased from Integrated DNA Technologies. UV-vis absorption spectroscopy was performed on an Agilent 8453 diode array spectrophotometer. Anaerobic procedures were carried out in a custom-designed MBraun glovebox equipped with a chiller at 15 °C. All solutions and proteins were made anaerobic on a Schlenk line by three cycles of evacuation (5 min) followed by flushing with Ar gas (10 min) before being brought into the glovebox. Nucleotides, *S*-adenosylmethionine (AdoMet), and NADPH were brought into the glovebox as lyophilized solids. *E. coli* TrxA and TrxB were purified according to published procedures (88, 89).

Cloning, Expression, and Purification of the NrdD and NrdG Proteins

The genes were amplified from the respective genomic DNA by PCR using Phusion polymerase (NEB) and were inserted into pET28a (Novagen), and linearized with NdeI and HindIII by Gibson isothermal assembly (90), yielding the plasmids pET28a-NrdD protein from *N. bacilliformis* (NbNrdD), pET28a-NrdG protein from *N. bacilliformis* (NbNrdG), and pET28a-NrdD from the deep-branching thermophilic bacterium *T. maritima* (TmNrdD). Because of the insolubility of the protein expressed from pET28a-NbNrdG, the *nrdG* gene was excised using the restriction enzymes NdeI and XhoI and ligated into pSV272, which contains a maltose-binding protein (MBP) with an N-terminal His₆-tag, and linearized with the same enzymes to obtain pMBP-NbNrdG. The plasmids pET28a-NbNrdD(C300A), pET28a-NbNrdD(C301A), and pET28a-NbNrdD(E438Q), containing the C300A, C301A, and E438Q mutations, respectively, were constructed by site-directed mutagenesis from pET28a-NbNrdD. All constructs were confirmed by DNA sequencing at the Massachusetts Institute of Technology Biopolymers Laboratory. The pET28a plasmid contains an N-terminal His₆ affinity purification tag followed by a thrombin cleavage site (MGSSH HHHHH SSSLV PRGSH-).

Expression and purification of the three proteins followed a similar protocol. The plasmids pET28a-TmNrdD, pET28a-NbNrdD, and pMBP-NbNrdG were separately transformed into BL21 (DE3) codon plus (RIL) cells (Stratagene) and grown on Luria-Bertani (LB)-agar plates with 50 µg/mL kanamycin (Kan). A single colony was inoculated into 5 mL starter culture of LB (50 µg/mL Kan in all growths), grown at 37 °C until saturated (12 h), and transferred into

200 mL of LB. Media for expression of NbNrdD and TmNrdD were supplemented with 50 μ M ZnSO₄. The cultures were grown at 37 °C with shaking at 200 rpm. At OD₆₀₀ ~0.6, the temperature was decreased to 20 °C, and isopropyl β -D-1-thiogalactopyranoside (Promega) was added to a final concentration of 0.1 mM. After 12 h, cells were pelleted by centrifugation (4,000 \times g, 10 min, 4 °C) and frozen at -80 °C. Typical yield was ~5 g of cell paste/L.

Cell paste (~1 g) was resuspended in 25 mL of 20 mM Tris, pH 7.5, 5 mM DTT, and 1 mM phenylmethanesulfonyl fluoride (buffer A). The cells were lysed by a single passage through a French pressure cell (14,000 psi). DNA was precipitated by dropwise addition of 5 mL of buffer A containing 6% (wt/vol) streptomycin sulfate. The mixture was shaken for an additional 10 min, and the precipitated DNA was removed by centrifugation (20,000 \times g, 10 min, 4 °C). Solid (NH₄)₂SO₄ was then added to 60% (wt/vol) saturation. The solution was shaken for an additional 20 min, and the precipitated protein was isolated by centrifugation (20,000 \times g, 10 min, 4 °C).

The pellet was dissolved in 30 mL of 20 mM Tris, pH 7.5, 5 mM *tris*(2-carboxyethyl)phosphine (TCEP), and 0.2 M KCl (buffer B) and incubated with 2 mL of TALON resin (Clontech) with shaking for 30 min. The column was then packed (0.8 \times 4 cm) and washed with 10 column volumes of buffer B, followed by 10 column volumes of buffer B containing 5 mM imidazole. Protein was eluted with 5 column volumes of buffer B containing 150 mM imidazole. The eluted protein was precipitated with solid (NH₄)₂SO₄ to 60% saturation and isolated by centrifugation (20,000 \times g, 10 min, 4 °C). The pellet was dissolved in 0.5 mL of buffer B and desalted using a Sephadex G-25 column (1.5 \times 8.5 cm, 15 mL), pre-equilibrated with 20 mM Tris, pH 7.5, 5% (vol/vol) glycerol, and 1 mM DTT (buffer C). The eluted protein was concentrated to ~200 μ M by ultrafiltration (Amicon YM-30), frozen in aliquots in liquid N₂, and stored at -80 °C. The final yield was 10 mg/g cells for NbNrdD (ϵ_{280} = 73,410 M⁻¹·cm⁻¹) and ~1 mg/g cells for MBP-NbNrdG (ϵ_{280} = 95,760 M⁻¹·cm⁻¹). SDS/PAGE gels of the purified proteins are shown in Figure V.13.

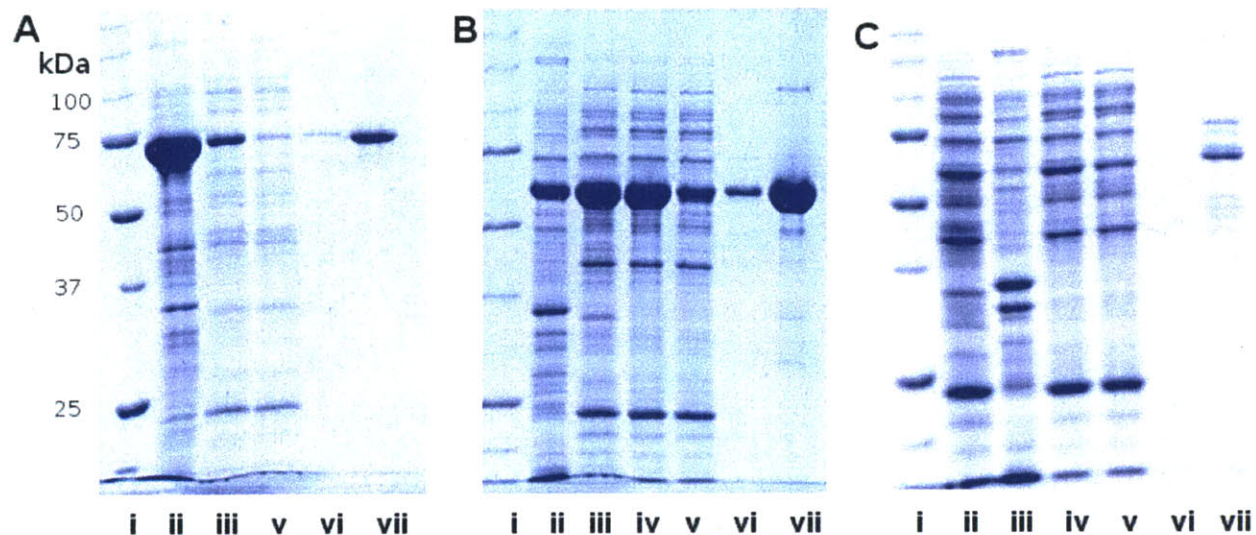


Figure V.13. Purification of NrdD and NrdG. (A) [SeMet]-TmNrdD; **(B)** NbNrdD, and **(C)** MBP-NbNrdG (SDS/PAGE 10% gels). Lanes: (1) ladder; (2) streptomycin sulfate precipitate; (3) streptomycin sulfate supernatant; (4) ammonium sulfate precipitate; (5) TALON column flow-through; (6) TALON column wash; and (7) TALON column eluted protein.

Reconstitution of the NbNrdG [4Fe4S] Cluster.

The procedure was carried out in a MBraun anaerobic chamber. Solutions of Na_2S and of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ in water (100 mM) were freshly prepared in the anaerobic chamber. A solution of MBP-NbNrdG (200 μM , 0.3 mL) was made anaerobic on a Schlenk line and brought into the glovebox. A solution of DTT (1 M) was added to 10 mM, followed by ordered addition of the solution of Na_2S (5 equivalents) and $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (5 equivalents). The mixture was incubated for 12 h at 4 °C. EDTA (5 eq.) was then added, and the solution was desalted using a Sephadex G-25 column (1 × 9 cm, 7 mL) equilibrated with Tris buffer (20 mM, pH 7.5). The final material typically contained ~2.5 atoms of Fe per peptide determined by the ferrozine assay (75).

Generation of the NbNrdD glycy radical.

In a 1.5-mL polypropylene Eppendorf tube, a 50- μL mixture of NbNrdD (40 μM), NbNrdG (20 μM), AdoMet (0.5 mM), Bicine potassium salt, pH 7.5 (30 mM), and acriflavin (10 μM) was placed 1 m away from a fluorescent lamp in the glovebox at 15 °C for 3 h. For inspection by X-band EPR spectroscopy, the solution was diluted to 200 μL with Tris buffer (20 mM, pH 7.5) and 5% (vol/vol) glycerol to give a final concentration of 10 μM NbNrdD and sealed in an EPR tube with a rubber stopper. The solution was quenched in liquid N_2 immediately after removal from the glovebox. The amount of glycy radical in the solution was determined by

comparing the EPR signal intensity to that of a CuSO_4 standard (76). A typical yield of 0.25–0.30 radicals per NbNrdD polypeptide was reproducibly obtained.

X-Band EPR Spectroscopy.

Continuous wave X-band EPR spectra were recorded at 77 K in the Massachusetts Institute of Technology Department of Chemistry Instrumentation Facility on a Bruker ESP-300 X-band spectrometer equipped with a quartz finger dewar filled with liquid N_2 . Experimental conditions were as follows: microwave frequency, 9.45 GHz; modulation amplitude, 0.15 mT; modulation frequency, 100 kHz; time constant, 5.12 ms; scan time, 41.9 s; microwave power, 20 μW .

Preparation of [^2H]-Gly-Labeled NbNrdD to Establish the Location of the Radical on Glycine.

The procedure was identical to the preparation of [SeMet]-TmNrdD, except that instead of SeMet, the culture contained L-methionine (50 mg/L) and [^2H]-glycine (6 mM, 98% isotopic enrichment; Cambridge Isotope Laboratories). The yield was ~1 g of cell paste, and purification was carried out according to the procedure described for the unlabeled protein. The EPR sample was then prepared as described above.

Assay for Solvent-Exchangeable Glycine H_α Protons.

Tris buffer (20 mM, pH 7.5) was prepared in D_2O (99.9%; Cambridge Isotope Laboratories) in the glovebox. NbNrdD and NbNrdG were exchanged into this buffer by repeated dilution and concentration by ultrafiltration (Amicon YM-30), such that <1% H_2O remained, and were incubated for 12 h at 4 °C to allow for proton exchange. The activation reaction was carried out as described above, but with all components made up in D_2O in the glovebox, followed by preparation of the EPR sample as described above.

Activity Assay for dCTP Formation by NbNrdD.

The assay mixture contained (in 100 μL) NbNrdD (4 μM , ~1 μM glycy radical), ATP (1 mM), 5- [^3H]-CTP (1 mM, 4,170 cpm/nmol), *E. coli* TrxA (30 μM), *E. coli* TrxB (1 μM), and NADPH (1 mM) in assay buffer (30 mM Tris, pH 7.5, 30 mM KCl, and 10 mM MgSO_4) and was incubated at 30 °C. Aliquots (20 μL) were removed at 2-min intervals and quenched with 2% (vol/vol) perchloric acid (20 μL). Subsequent to removal of the phosphates using calf intestine alkaline phosphatase (Roche), dCTP formation was analyzed by the method of Steeper and

Steuart (77). One unit of activity is equivalent to 1 nMol of dCTP per minute. The specific activity of NbNrdD is 49 U/mg NrdD protein ($\sim 0.24 \text{ s}^{-1}$ per glycy radical).

Stoichiometry of NADPH Consumption and dCTP Production.

The assay mixture was divided into 20- μL aliquots containing NbNrdD (4 μM , $\sim 1 \mu\text{M}$ glycy radical), dATP (0.1 mM), 5- $[\text{}^3\text{H}]$ -CTP (1 mM, 4170 cpm/nmol), *E. coli* TrxA (5 μM), *E. coli* TrxB (1 μM), and NADPH (0, 70, 140, or 210 μM) in assay buffer and was incubated at 30 °C for 3 h to allow for complete consumption of NADPH. Workup of the samples was carried out as described above to quantify the dCTP formed.

Activity Assay for Cytosine Release by NbNrdD(C300A).

The assay mixture contained (in 100 μL) NbNrdD(C300A) (33 μM , $\sim 8 \mu\text{M}$ glycy radical) and 5- $[\text{}^3\text{H}]$ -CTP in assay buffer and was incubated at 30 °C. Aliquots (20 μL) were removed at 2, 4, 8, and 16 min and quenched with 2% (vol/vol) perchloric acid (20 μL). dCTP formation was analyzed by the method of Steeper and Steuart (77). Formation of Cyt was analyzed by passing the mixture through an anion exchange column to remove the nucleoside triphosphates as previously described (31).

To a 7-mL portion of the eluate of the Dowex-1-borate column was added carrier Cyt and dC (10 nMol each). The mixture was concentrated by lyophilization, redissolved in water, and cooled on ice, and the precipitated borate salts were removed by centrifugation. The supernatant was analyzed by HPLC using an Alltech Econosil column (C_{18} , 10 μM , $250 \times 4.6 \text{ mM}$) on a Waters 515 HPLC system equipped with a 2,996 photodiode array detector. The compounds were eluted with potassium phosphate buffer (20 mM, pH 6.8) at a flow rate of 1.0 mL/min. Fractions were collected (0.5 mL) and analyzed by scintillation counting. Cyt was identified by coelution with a standard at 5 min (Figure V.7).

Preparation of [SeMet]-Labeled TmNrdD for Crystallography.

The procedure was adapted from existing protocols (74). A 5 mL starter culture was prepared according to the procedure described above, harvested by centrifugation ($3,000 \times g$, 10 min, 4 °C), and transferred into 200 mL of M9 minimal medium supplemented with glucose (0.4%), thiamine (50 mg/L), Kan (50 mg/L), ZnSO_4 (50 μM), FeCl_3 (10 μM), MgCl_2 (2 mM), CaCl_2 (0.1 mM), and the L-amino acids lysine (100 mg/L), phenylalanine (100 mg/L), threonine (100 mg/L), isoleucine (50 mg/L), leucine (50 mg/L), and valine (50 mg/L). The culture was

grown at 37 °C with shaking at 220 rpm. At $OD_{600} \sim 0.3$, SeMet (50 mg/L) was added, followed by shaking for 20 min. The temperature was lowered to 20 °C, and IPTG (Promega) was added to a final concentration of 0.2 mM. After 12 h, cells were pelleted by centrifugation ($4,000 \times g$, 10 min, 4 °C) and frozen at -80 °C. The yield was ~ 1 g of cell paste, and purification was carried out according to the procedure described for NbNrdD. The final yield was ~ 2 mg/g cells ($\epsilon_{280} = 106,830 \text{ M}^{-1} \cdot \text{cm}^{-1}$). An SDS/PAGE gel of the purified protein is shown in Figure V.13.

Crystallization and Crystal Structure of TmNrdD.

Crystals of SeMet-TmNrdD were grown aerobically by sitting drop vapor diffusion at 21 °C. Protein, with the His-tag intact, at 13 mg/mL in a buffer containing 25 mM HEPES, pH 7.6, 15 mM MgCl_2 , 20 mM KCl, 0.5 mM TCEP, 1 mM dGTP, and 5 mM ATP, was screened against commercial screens (Hampton Research, Microlytic, and Qiagen) at a 1:1 ratio of protein to precipitant. A Phoenix pipetting robot (Art Robbins Instruments) was used for dispensing 150-nL drops in screening trays. Diffraction quality crystals grew over several weeks in wells containing 0.085 M trisodium citrate, pH 5.6, 0.17 M ammonium acetate, 25.5% (wt/vol) PEG 4000, and 15% (vol/vol) glycerol. Crystals were flash frozen in liquid nitrogen without additional cryoprotection.

The structure of TmNrdD was solved by single-wavelength anomalous dispersion. A 1.64-Å resolution Se peak anomalous dataset (12664.1 eV) was collected at the Advanced Photon Source using a minikappa goniometer to collect Friedel pairs on a single image. Data were collected on a Pilatus 6M detector (Dectris). The data were indexed and scaled with HKL2000 (78) in space group P21 with unit cell constants $a = 78.3$, $b = 98.8$, $c = 86.6$, and $\beta = 111.7$. The resulting unit cell volume of $622,000 \text{ \AA}^3$ is consistent with two molecules per asymmetric unit with a solvent content of $\sim 40\%$ (statistics found in Table V.2). Forty-five initial Se sites were found with SHELXD in the package HKL2MAP (79) with a resolution cutoff of 2.1 Å ($d''/\text{sig} = 0.8$). The resulting sites were refined using data to 1.8-Å resolution, and density modification was performed in Phenix AutoSol (80), yielding the phasing statistics shown in Table V.2. The resulting maps at 1.8-Å resolution were adequate for chain tracing and manual building of the entire protein model and surrounding water molecules in Coot (81). The resolution was increased to 1.64 Å, and iterative refinement of the model was performed with phenix.refine. Composite omit maps were used to verify residue and ligand placement. All native residues are present at the N and C termini of chain A; chain B is missing three C-terminal residues. Both chains have

additional density at the N terminus corresponding to eight residues of the thrombin cleavage site and linker; no residues of the His-tag are observed, although the His-tag was not cleaved before crystallization. Residues 54–62 of chain A are unstructured. Residues 330–349 of chain A and 328–350 of chain B are not observed in the density at all as a consequence of peptide bond cleavage near residue 330 (Figure V.10). The final model contained 98.3% of residues in the favored region of the Ramachandran plot, with 1.6% in additionally allowed regions and 0.08% (1 residue, G621) disallowed. G621 is converted in to the glycy radical and thus is expected to adopt a slightly strained conformation in the unactivated protein. Figures were created in PyMOL Version 1.4.1 (Schrödinger, LLC). A structural model for TmNrdD Cys loop was constructed by superimposing the structures of T4 bacteriophage (1H79) (45) and TmNrdD using residues in the β -barrel surrounding the active site loops. Because there are no structures of a nucleoside triphosphate-bound RNR, CTP was modeled in the active site based on the position of CDP in the class II TmNrdJ (1XNJ) (46). The resulting hybrid model should be considered purely as a tool for guiding discussions of the possible chemistry occurring in TmNrdD and was not subjected to energy minimizations or dynamics.

(figure next page)

Table V.2 Data collection and refinement statistics

	SeMet TmNrdD
Data collection and processing¹	
Dataset	Se peak
Resolution (Å)	50-1.64 (1.67-1.64)
Completeness %	94.4 (65.4)
$\langle I/\sigma \rangle$	6.6 (1.1) ²
R _{sym}	0.059 (0.64)
Unique reflections	279509 (9748)
Redundancy	1.8 (1.6)
CC1/2	(0.55)
Phasing	
Phasing resolution	50-1.80
Figure of merit (FOM)	0.42
Density modification FOM	0.67
Density modification R factor	0.25
Model refinement	
Resolution (Å)	50-1.64
R _{work} , R _{free} ³	0.171, 0.200
Protein atoms	10331
Water molecules	892
RMS ⁴ Bond length deviation (Å)	0.050
RMS Bond angle deviation (°)	0.928
Rotamer outliers (%)	0.90
Average B factors (Å ²)	
Monomer A	21.6
Monomer B	20.6
Water	35.5

¹Statistics were calculated with Friedel pairs kept separate during scaling. Numbers in parentheses are for highest resolution bin.

²Data were trimmed based on CC1/2 value rather than the $\langle I/\sigma \rangle$ value. Inclusion of highest resolution data improved the quality of the electron density maps

³3783 reflections were set aside prior to refinement for cross validation

⁴RMS, root mean squared

Phylogenetic Analysis of NrdDs.

To determine the existence of different NrdD subtypes, 59 amino acid sequences were chosen to include phylogenetically and metabolically diverse bacteria and archaea and aligned with Muscle (MEGA5 software) (82). Subsequently, a neighbor-joining phylogenetic tree (83) was generated, and the Jones-Taylor-Thornton matrix method (84) was used to determine the evolutionary distances. The unequal rate of variation among amino acid sites was modeled with a γ distribution with shape parameter of 4 (85). Additionally, a complete deletion of the sites containing gaps was used to palliate the length and divergence variation, focusing on more conserved amino acid clusters. The level of confidence for the branches was determined based on 2,000 bootstrap replicates (86). The resulting consensus tree was rendered using the web-based program iTOL (87) (Figure V.12B, bootstrap values presented in Figure V.14).

(figure next page)

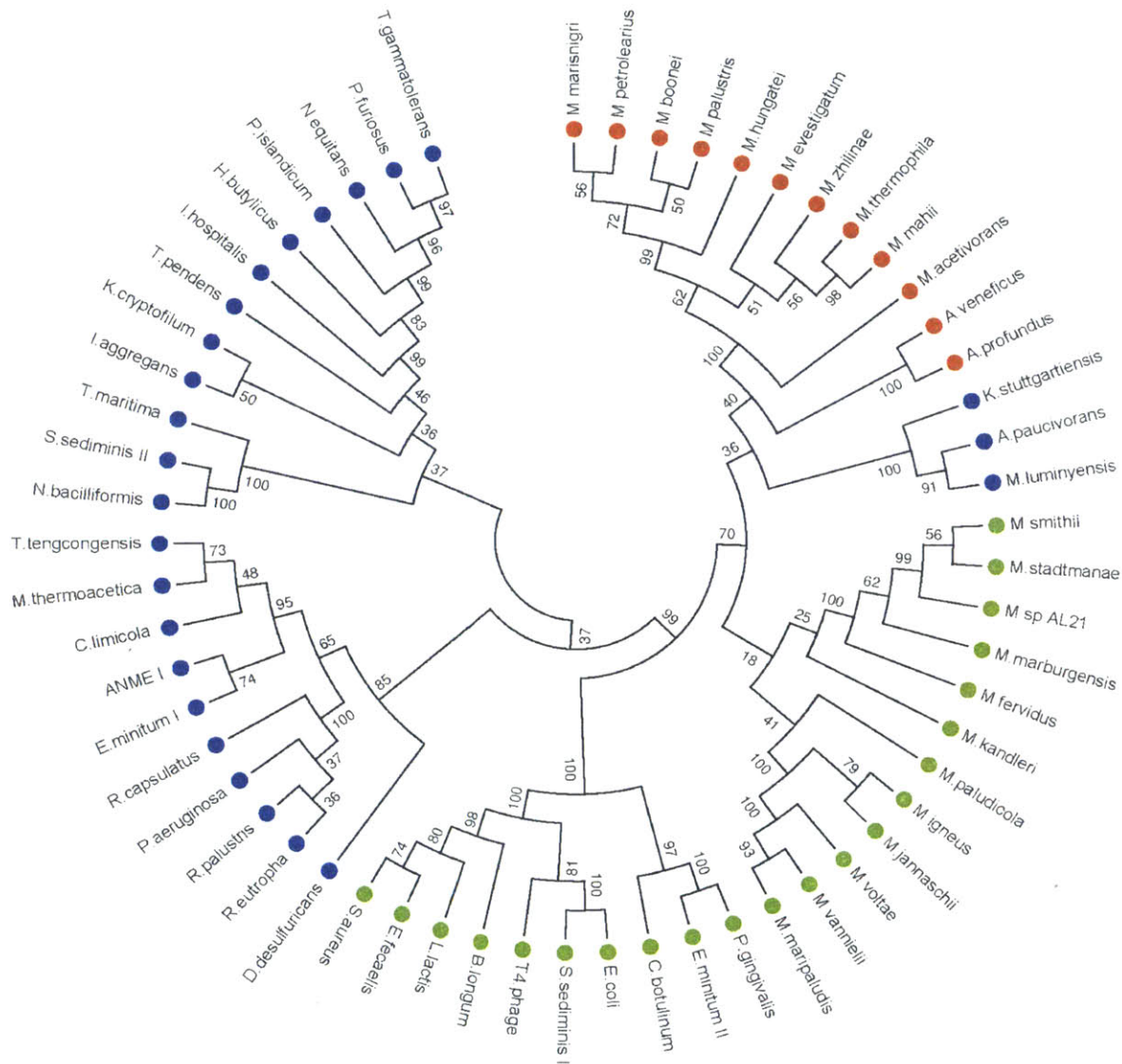


Figure V.14. Unrooted phylogenetic tree of NrdDs as shown in Figure V.12B. Bootstrap values resulting from 2,000 resamplings are given for each branch. Nrd1, Nrd2, and Nrd3 are labeled green, blue, and red, respectively.

IV.F Acknowledgements

We thank Dr. Xiang-Yang Han for the gift of *N. bacilliformis* genomic DNA, Prof. Kenneth Noll for the gift of *T. maritima* genomic DNA, and Dr. Greg Fournier, Dr. Silvan Scheller, and Prof. Rudolf K. Thauer for helpful discussions. This work was supported in part by ASTAR Singapore (Y.W.) and the National Science Foundation Graduate Research Fellowship under Grant 0645960 (M.A.F.). C.L.D. is a Howard Hughes Medical Institute Investigator. This

work is based on research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by National Institutes of Health (NIH) Award GM103403 from the National Center for Research Resources. Use of the Advanced Photon Source is supported by the US Department of Energy, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. This work was supported by NIH Grant GM29595 (J.S.).

IV.G References

1. Fontecave M, Eliasson R, Reichard P (1989) Oxygen-sensitive ribonucleoside triphosphate reductase is present in anaerobic *Escherichia coli*. *Proc Natl Acad Sci USA* 86(7):2147–2151.
2. Sun X, et al. (1996) The free radical of the anaerobic ribonucleotide reductase from *Escherichia coli* is at glycine 681. *J Biol Chem* 271(12):6827–6831.
3. Nordlund P, Reichard P (2006) Ribonucleotide reductases. *Annu Rev Biochem* 75: 681–706.
4. Mulliez E, Ollagnier S, Fontecave M, Eliasson R, Reichard P (1995) Formate is the hydrogen donor for the anaerobic ribonucleotide reductase from *Escherichia coli*. *Proc Natl Acad Sci USA* 92(19):8759–8762.
5. Hofer A, Crona M, Logan DT, Sjöberg B-M (2012) DNA building blocks: Keeping control of manufacture. *Crit Rev Biochem Mol Biol* 47(1):50–63.
6. Stubbe J (1998) Ribonucleotide reductases in the twenty-first century. *Proc Natl Acad Sci USA* 95(6):2723–2724.
7. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: Functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* 29(5):1097–1106.
8. Gambarelli S, Luttringer F, Padovani D, Mulliez E, Fontecave M (2005) Activation of the anaerobic ribonucleotide reductase by S-adenosylmethionine. *ChemBioChem* 6(11):1960–1962.
9. Holmgren A (1979) Glutathione-dependent synthesis of deoxyribonucleotides. Characterization of the enzymatic mechanism of *Escherichia coli* glutaredoxin. *J Biol Chem* 254(9):3672–3678.

10. Jordan A, Åslund F, Pontis E, Reichard P, Holmgren A (1997) Characterization of *Escherichia coli* NrdH: A glutaredoxin-like protein with a thioredoxin-like activity profile. *J Biol Chem* 272(29):18044–18050.
11. Avval FZ, Holmgren A (2009) Molecular mechanisms of thioredoxin and glutaredoxin as hydrogen donors for mammalian S phase ribonucleotide reductase. *J Biol Chem* 284(13):8233–8240.
12. Young P, Ohman M, Xu MQ, Shub DA, Sjöberg BM (1994) Intron-containing T4 bacteriophage gene sunY encodes an anaerobic ribonucleotide reductase. *J Biol Chem* 269(32):20229–20232.
13. Torrents E, et al. (2000) The anaerobic (class III) ribonucleotide reductase from *Lactococcus lactis*: Catalytic properties and allosteric regulation of the pure enzyme system. *J Biol Chem* 275(4):2463–2471.
14. Sawers G, Böck A (1988) Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *J Bacteriol* 170(11):5330–5336.
15. Knappe J, Sawers G (1990) A radical-chemical route to acetyl-CoA: The anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol Lett* 75(4): 383–398.
16. Axley MJ, Grahame DA (1991) Kinetics for formate dehydrogenase of *Escherichia coli* formate-hydrogenlyase. *J Biol Chem* 266(21):13731–13736.
17. Torrents E, Aloy P, Gibert I, Rodríguez-Trelles F (2002) Ribonucleotide reductases: Divergent evolution of an ancient enzyme. *J Mol Evol* 55(2):138–152.
18. Lundin D, Gribaldo S, Torrents E, Sjöberg B-M, Poole AM (2010) Ribonucleotide reduction-horizontal transfer of a required function spans all three domains. *BMC Evol Biol* 10(1):383.
19. Kim BH, Gadd GM (2008) *Bacterial Physiology and Metabolism* (Cambridge Univ Press, Cambridge, UK).
20. Licht S, Stubbe J (1999) *Comprehensive Natural Products Chemistry*, eds Barton S, Nakanishi K, Meth-Cohn O, Poulter C (Elsevier Science, New York), Vol 5, pp 163.
21. Eklund H, Fontecave M (1999) Glycyl radical enzymes: A conservative structural basis for radicals. *Structure* 7(11):R257–R262.

22. Stubbe J, Ackles D (1980) On the mechanism of ribonucleoside diphosphate reductase from *Escherichia coli*. Evidence for 3'-C-H bond cleavage. *J Biol Chem* 255(17): 8027–8030.
23. Stubbe J, Ackles D, Segal R, Blakley RL (1981) On the mechanism of ribonucleoside triphosphate reductase from *Lactobacillus leichmannii*. Evidence for 3' C-H bond cleavage. *J Biol Chem* 256(10):4843–4846.
24. Stubbe J, Ator M, Krenitsky T (1983) Mechanism of ribonucleoside diphosphate reductase from *Escherichia coli*. Evidence for 3'-C-H bond cleavage. *J Biol Chem* 258(3): 1625–1631.
25. Licht S, Gerfen GJ, Stubbe J (1996) Thiyl radicals in ribonucleotide reductases. *Science* 271(5248):477–481.
26. Andersson J, Bodevin S, Westman M, Sahlin M, Sjöberg B-M (2001) Two active site asparagines are essential for the reaction mechanism of the class III anaerobic ribonucleotide reductase from bacteriophage T4. *J Biol Chem* 276(44):40457–40463.
27. Mao SS, et al. (1992) A model for the role of multiple cysteine residues involved in ribonucleotide reduction: Amazing and still confusing. *Biochemistry* 31(40):9733–9743.
28. Lawrence CC, et al. (1999) High-field EPR detection of a disulfide radical anion in the reduction of cytidine 5'-diphosphate by the E441Q R1 mutant of *Escherichia coli* ribonucleotide reductase. *Proc Natl Acad Sci USA* 96(16):8979–8984.
29. Logan DT, Andersson J, Sjöberg B-M, Nordlund P (1999) A glyceryl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* 283(5407):1499–1504.
30. Andersson J, Westman M, Sahlin M, Sjöberg B-M (2000) Cysteines involved in radical generation and catalysis of class III anaerobic ribonucleotide reductase: A protein engineering study of bacteriophage T4 NrdD. *J Biol Chem* 275(26):19449–19455.
31. Wei Y, et al. (2014) A chemically competent thiosulfuranyl radical on the *Escherichia coli* class III ribonucleotide reductase. *J Am Chem Soc* 136(25):9001–9013.
32. Lenz R, Giese B (1997) Studies on the mechanism of ribonucleotide reductases. *J Am Chem Soc* 119(12):2784–2794.
33. Lundin D, Torrents E, Poole A, Sjöberg B-M (2009) RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. *BMC Genomics* 10(1):589.

34. Han XY, Hong T, Falsen E (2006) *Neisseria bacilliformis* sp. nov. isolated from human infections. *J Clin Microbiol* 44(2):474–479.
35. Masliah-Planchon J, et al. (2009) Endocarditis due to *Neisseria bacilliformis* in a patient with a bicuspid aortic valve. *J Clin Microbiol* 47(6):1973–1975.
36. Wagner AF, Frey M, Neugebauer FA, Schäfer W, Knappe J (1992) The free radical in pyruvate formate-lyase is located on glycine-734. *Proc Natl Acad Sci USA* 89(3):996–1000.
37. Parast CV, et al. (1995) Hydrogen exchange of the glycy radical of pyruvate formate-lyase is catalyzed by cysteine 419. *Biochemistry* 34(8):2393–2399.
38. Mulliez E, Fontecave M, Gaillard J, Reichard P (1993) An iron-sulfur center and a free radical in the active anaerobic ribonucleotide reductase of *Escherichia coli*. *J Biol Chem* 268(4):2296–2299.
39. Laurent TC, Moore EC, Reichard P (1964) Enzymatic synthesis of deoxyribonucleotides
40. IV. Isolation and characterization of thioredoxin, the hydrogen donor from *Escherichia coli* B. *J Biol Chem* 239(10):3436–3444.
40. Moore EC, Reichard P, Thelander L (1964) Enzymatic synthesis of deoxyribonucleotides V. Purification and properties of thioredoxin reductase from *Escherichia coli* B. *J Biol Chem* 239(10):3445–3452.
41. Goulian M, Beck WS (1966) Purification and properties of cobamide-dependent ribonucleotide reductase from *Lactobacillus leichmannii*. *J Biol Chem* 241(18):4233–4242.
42. Reichard P (2002) Ribonucleotide reductases: The evolution of allosteric regulation. *Arch Biochem Biophys* 397(2):149–155.
43. Luttringer F, Mulliez E, Dublet B, Lemaire D, Fontecave M (2009) The Zn center of the anaerobic ribonucleotide reductase from *Escherichia coli*. *J Biol Inorg Chem* 14(6):923–933.
44. Ge J, Yu G, Ator MA, Stubbe J (2003) Pre-steady-state and steady-state kinetic analysis of *Escherichia coli* class I ribonucleotide reductase. *Biochemistry* 42(34):10071–10083.
45. Larsson K-M, Andersson J, Sjöberg B-M, Nordlund P, Logan DT (2001) Structural basis for allosteric substrate specificity regulation in anaerobic ribonucleotide reductases. *Structure* 9(8):739–750.
46. Larsson K-M, et al. (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase. *Nat Struct Mol Biol* 11(11):1142–1149.

47. Persson AL, Sahlin M, Sjöberg B-M (1998) Cysteinylyl and substrate radical formation in active site mutant E441Q of *Escherichia coli* class I ribonucleotide reductase. *J Biol Chem* 273(47):31016–31020.
48. Sintchak MD, Arjara G, Kellogg BA, Stubbe J, Drennan CL (2002) The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nat Struct Biol* 9(4):293–300.
49. Vey JL, et al. (2008) Structural basis for glycylyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137–16141.
50. Peng Y, Veneziano SE, Gillispie GD, Broderick JB (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J Biol Chem* 285(35):27224–27231.
51. Funk MA, Judd ET, Marsh ENG, Elliott SJ, Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycylyl radical enzyme activation and activity. *Proc Natl Acad Sci USA* 111(28):10161–10166.
52. Wood GE, Haydock AK, Leigh JA (2003) Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*. *J Bacteriol* 185(8):2548–2554.
53. Thauer R, Rupprecht E, Jungermann K (1970) The synthesis of one-carbon units from CO₂ via a new ferredoxin dependent monocarboxylic acid cycle. *FEBS Lett* 8(5):304–307.
54. Thauer RK, Kirchniawy FH, Jungermann KA (1972) Properties and function of the pyruvate-formate-lyase reaction in Clostridia. *Eur J Biochem* 27(2):282–290.
55. Poehlein A, et al. (2012) An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PLoS ONE* 7(3):e33439.
56. Sun H, et al. (2010) Complete genome sequence of *Desulfarculus baarsii* type strain (2st14). *Stand Genomic Sci* 3(3):276–284.
57. Gorlas A, Robert C, Gimenez G, Drancourt M, Raoult D (2012) Complete genome sequence of *Methanomassiliicoccus luminyensis*, the largest genome of a human-associated Archaea species. *J Bacteriol* 194(17):4745.
58. Klenk H-P, et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390(6658):364–370.

59. Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M (2013) A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol Biol* 13(1):33.
60. Poole AM, Logan DT, Sjöberg B-M (2002) The evolution of the ribonucleotide reductases: Much ado about oxygen. *J Mol Evol* 55(2):180–196.
61. Cotruvo JA, Jr, Stubbe J (2012) Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: The class I ribonucleotide reductases as a case study. *Metallomics* 4(10):1020–1036.
62. Herlemann DPR, et al. (2009) Genomic analysis of “*Elusimicrobium minutum*,” the first cultivated representative of the phylum “Elusimicrobia” (formerly termite group 1). *Appl Environ Microbiol* 75(9):2841–2849.
63. Ragsdale SW, Pierce E (2008) Acetogenesis and the Wood–Ljungdahl pathway of CO₂ fixation. *Biochim Biophys Acta* 1784(12):1873–1898.
64. Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R (2008) Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 6(8):579–591.
65. White RH (1997) Purine biosynthesis in the domain Archaea without folates or modified folates. *J Bacteriol* 179(10):3374–3377.
66. Fricke WF, et al. (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *J Bacteriol* 188(2):642–658.
67. Sakai S, et al. (2011) Genome sequence of a mesophilic hydrogenotrophic methanogen *Methanocella paludicola*, the first cultivated representative of the order Methanocellales. *PLoS ONE* 6(7):e22898.
68. Maeder DL, et al. (2006) The *Methanosarcina barkeri* genome: Comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within Methanosarcinal genomes. *J Bacteriol* 188(22):7922–7931.
69. Allen MA, et al. (2009) The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: The role of genome evolution in cold adaptation. *ISME J* 3(9): 1012–1035.
70. Michaelis W, et al. (2002) Microbial reefs in the Black sea fueled by anaerobic oxidation of methane. *Science* 297(5583):1013–1015.

71. Milucka J, et al. (2012) Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* 491(7425):541–546.
72. Reichard P (1993) From RNA to DNA, why so many ribonucleotide reductases? *Science* 260(5115):1773–1777.
73. Stubbe J (2000) Ribonucleotide reductases: The link between an RNA and a DNA world? *Curr Opin Struct Biol* 10(6):731–736.
74. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229(1):105–124.
75. Fish W (1988) Rapid colorimetric micromethod for the quantitation of complexed iron in biological samples. *Methods Enzymol* 158:357.
76. Malmström BG, Reinhammar B, Vänngård T (1970) The state of copper in stellacyanin and laccase from the lacquer tree *Rhus vernicifera*. *Biochim Biophys Acta* 205(1):48–57.
77. Steeper J, Steuart C (1970) A rapid assay for CDP reductase activity in mammalian cell extracts. *Anal Biochem* 34(1):123–130.
78. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymology*, ed Charles W. Carter, Jr. (Academic Press, New York), Vol 276, pp 307–326.
79. Pape T, Schneider TR (2004) HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J Appl Cryst* 37(5):843–844.
80. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
81. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
82. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
83. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425.
84. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275–282.

85. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8(8):610–618.
86. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Biol* 27(4):401–410.
87. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127–128.
88. Russel M, Model P (1985) Direct cloning of the *trxB* gene that encodes thioredoxin reductase. *J Bacteriol* 163(1):238–242.
89. Chivers PT, et al. (1997) Microscopic pKa values of *Escherichia coli* thioredoxin. *Biochemistry* 36(48):14985–14991.
90. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
91. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1):539.

Chapter VI

Recognizing the diversity within the glycy radical enzyme family

VI.A Summary

Known glycyl radical enzymes (GREs) catalyze a diverse range of reactions that include C–C, C–O, and C–N cleavage. These enzymes are broadly distributed within anaerobic bacteria and contribute to metabolism by initiating radical chemistry on difficult substrates. Current genomic sequencing efforts have revealed a large number of GRE sequences of unknown function, but the breadth of the chemistry that these enzymes catalyze, as well as how they control substrate specificity, has not yet been fully documented. The structures and biochemical results presented in previous chapters of this thesis have sparked a number of questions about GRE evolution, adaptability, and relation to radical enzymes that perform similar chemistry. In particular, Chapters III and IV revealed active site restructuring that enables binding and activation of specific substrates within the common GRE active site architecture. Within this chapter, I will analyze a diverse group of GRE sequences and use new bioinformatics tools to clarify the scope of the puzzle that these enzymes pose.

VI.B Unexplored function of glycyl radical enzyme family members

Although it has been known since at least the 1990s that GREs in the ribonucleotide reductase (RNR)/pyruvate formate-lyase (PFL) superfamily catalyze numerous and diverse reactions (1, 2), attempts to count or categorize individual subgroups of enzymes that are isofunctional have been hampered by inadequate structural and biochemical characterization. This effort is further complicated by the fact that GREs tend to be fairly similar—for example, the two eliminases choline trimethylamine-lyase (CutC) and glycerol dehydratase (GDH) are ~36% identical despite working on completely different substrates, and benzylsuccinate synthase (BSS) and GDH are ~29% identical despite catalyzing totally different reactions with different substrates. This similarity in part stems from the fact that all GREs share the core 10-stranded β/α barrel architecture, and thus much of the protein is conserved to retain basic structural features. A second shared feature is that all GRE reactions have a common set of activation and initiation steps (see Section I.G), which may constrain evolution of residues in regions of the protein other than the active site. Another explanation for the similarity of GREs, especially those in the dehydratase group including CutC and GDH, is that these enzymes are evolving rapidly to accommodate new substrates for the same general reaction, a 1,2-elimination. Gene duplication has enabled an expansion of GRE functions, and subsequent selection for new

functions would favor changes to residues that contribute directly to binding and activation of new substrates rather than the enzyme as a whole. Since CutC and GDH contribute as branches into central metabolism, they may also be under less selection to preserve enzymatic activity than a central metabolic enzyme like PFL, which is at the intersection of several important catabolic and anabolic pathways. When combined with lateral gene transfer (LGT), movement of genetic information from one bacterial species to an unrelated species, the result is a group of rapidly evolving and widely distributed enzymes that have adapted to provide new metabolic functions, often based on pre-existing pathways.

Previous attempts to cluster GREs by phylogeny identified several subgroupings based on enzymes of known function and suggested that active site conservation was correlated between these subgroups (1). The active site of GREs is composed of a 10-stranded β/α barrel with a central pair of loops (Gly loop and Cys loop) that contain the catalytic machinery for generation of a substrate-based radical. The strands of the barrel ($\beta 1$ - $\beta 10$) contribute residues to the central active site that appear to be important for determining the course of the radical reaction as well as the specific substrate in a number of structurally and biochemically characterized systems. However, there is still a dearth of structural information within this family, and previous efforts to understand GRE chemistry have focused on the Cys loop, which harbors a conserved cysteine (colored red in this section) that is thought to be converted to a thiyl radical during catalysis; the thiyl radical serves as the initiator of radical chemistry. A number of sequence motifs within this loop have been identified that give clues as to function (1). For example, the CCVS motif in PFL, in which a pair of adjacent cysteines facilitate pyruvate cleavage to formate with acetylation of one of the cysteines (underlined); or the GCVE motif in GDH, where a glutamate two residues away (underlined) serves as a catalytic base during elimination of water from glycerol (3). However, this strategy is somewhat limited as enzymes that perform similar chemistry on different substrates may use similar pieces within the active site to catalyze that chemistry. For example, CutC also contains the GCVE motif even though CutC has no glycerol dehydratase activity. This similarity in CutC and GDH is not entirely surprising as these enzymes are closely related; however, the very distantly related class I and II RNRs also contain a CXE motif that places the Cys loop glutamate adjacent to the substrate so it can act as a general base. The structure of these loops is somewhat different, and it appears that they may have arisen through convergent evolution, as neither class III RNRs nor PFL contain

this Cys loop glutamate. Another pitfall of sequence motifs is that they can be retained in enzymes performing different chemical reactions. For example, *p*-hydroxyphenylacetate decarboxylase (HPD) contains the same GCVE motif used to do elimination chemistry in GDH and CutC, but HPD catalyzes decarboxylation rather than 1,2-elimination. Clearly, a more nuanced strategy that takes into account structural information, microbiology, and genomics is necessary to unravel the range of biochemical functions performed by GREs.

In order to better grasp the diversity and interconnections within the GRE family, I analyzed sequences within the InterPro family IPR004184, which comprises PFL and its relatives, including the structurally characterized members GDH, CutC, BSS, and HPD. InterPro is a curated consortium of databases that is organized around enzyme function (2). The analysis was performed using the Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST) (3), which uses BLAST (Basic Local Alignment Search Tool) to generate all-by-all similarity scores between a large group of sequences. The data are visualized as a sequence similarity network (Figure VI.1) (4). This diagram depicts clusters of sequences as nodes (filled squares) interconnected by edges (curved lines). The length of the lines is roughly inversely proportional to the similarity and interconnectedness of the node such that highly connected or highly similar nodes cluster together and less similar nodes are distantly spaced. The utility of these diagrams is visualization of diverse networks of sequences that do not have a clear evolutionary history, such as with groups of enzymes that have undergone duplication and divergence (paralogs), as GREs clearly are. At a low sequence identity threshold (as in Figure VI.1), relationships are visible between distantly related sequence clusters, whereas increasing the percent identity threshold in steps reveals a more granular picture of sequence relationships. In contrast, tree diagrams are useful in phylogenetic analyses to show ancestry and vertical inheritance of true orthologs, but can be misleading when sequences and sequence clusters have undergone rapid divergence with few intermediate sequences. Class III RNRs and YjII proteins—another group of distantly related, putative GREs that have not been characterized—are not included in the analysis currently, as they are beyond the scope of this section. The broader family will be discussed more below in Section VI.G. For the purposes of this section, I will restrict the meaning of GREs to those under consideration in Figure VI.1.

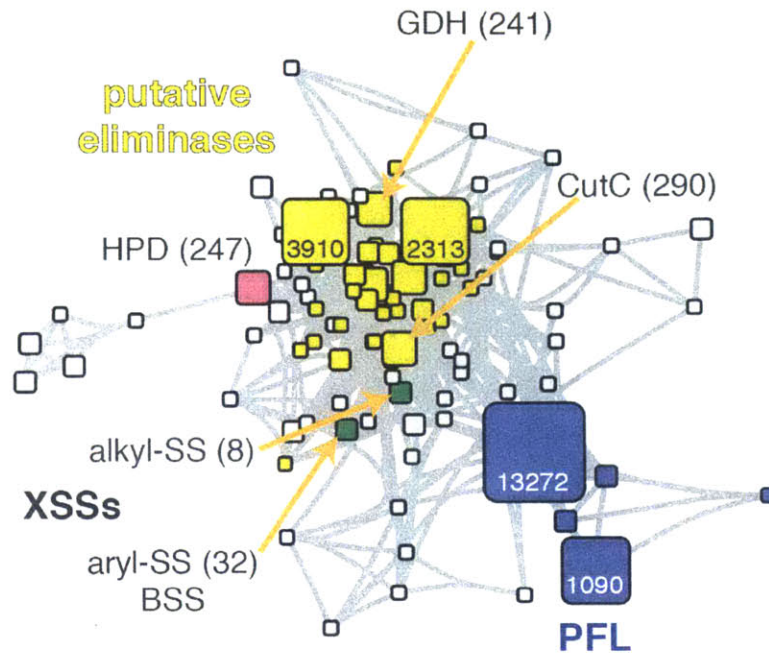


Figure VI.1. Low threshold sequence similarity network of GREs. Nodes (squares) represent groups of sequences clustered at 40% identity, a threshold at which isofunctional genes should cluster together in just a few representative nodes. Nodes are sized based on number of sequences (not directly proportional) with larger nodes representing more sequences; the top four nodes are labeled with the number of included sequences. Nodes containing characterized enzymes are noted with number of included sequences given in parentheses. Edges (gray connecting lines) are drawn between nodes that are at least 28.5% identical, the highest value for which all nodes remain connected. A total of 97 nodes are shown, representing 27,470 sequences. Nodes are classified broadly into PFL (blue) and putative eliminase (yellow) classes (including GDH and CutC), with known members highlighted. These designations are based on structures and biochemical data and are consistent with conservation of active site motifs (see text). The clusters were defined at an identity threshold of 32%. Two classes of BSS-related X-succinate synthase (SS) enzymes (XSSs, green) and HPD (magenta) are shown. There are 4 putative eliminase nodes of unknown function with more than 200 members, and 6 nodes with 20-200 members. There are 46 nodes (gray) outside of the proposed groupings.

Within the GRE family, PFL sequences represent more than half the total number of sequences within the database, but they are distributed in only a few nodes, suggesting very high conservation within PFL sequences. In contrast, other GREs, including the putative eliminases, are widely dispersed across the network over many nodes, indicating that these sequences have much lower overall identity. The putative eliminases are tightly clustered, suggesting that they have a common origin and have diverged less than many of the outlier sequences. The putative eliminases are also present near the center of the network, which implies that they represent a

branching point from which other sequences have diverged (discussed in Section VI.E). Interestingly, X-succinate synthases (XSSs) including BSS and HPD do not appear to be as related to each other as to members of the putative eliminase group, despite the fact that these two non-eliminases bind similar substrates/products and both contain small, [4Fe-4S]-cluster-binding proteins. It thus appears that the ability to bind an aromatic substrate and the acquisition of accessory subunits occurred through convergent, rather than divergent, evolution for these two enzymes. This conclusion is supported by the comparison of the subunit architecture of these enzymes presented in Chapter II, which shows a completely different binding mode and function for the small subunits. The substrate-bound structure of BSS, presented in Chapter III, revealed that the binding site for hydrocarbons is along a hydrophobic wall adjacent to the binding site for fumarate. The binding site for 4-hydroxyphenylacetate in HPD is quite different from the binding site for toluene observed in BSS, despite the similarity of these substrates, further supporting the conclusion that these enzymes are not directly related.

To gain insight into the function of uncharacterized GREs, I constructed another sequence similarity network with a more stringent identity threshold for nodes and edges. As noted above, the GRE family is broadly split between PFLs, putative eliminases, and a diverse group of outliers that contains at least two enzymes with known non-eliminase function, BSS and HPD. Within these putative non-eliminases, which do not represent a distinct cluster within Figure VI.1, but instead are outliers from the PFL and eliminase clusters, there are eight large sequence clusters and many more single nodes that are only distantly related to each other or other GREs (Figure VI.2A). Only two of these clusters contain enzymes that have been structurally and/or biochemically characterized. The XSS cluster is split into several groups that have been annotated according known substrates (see figure legend). As expected from the structural analysis presented in Chapter III, the major division is between aryl-SSs and alkyl-SSs. Interestingly, it is clear from the node arrangement that 2-naphthylmethyl-SS and 1-methylalkyl-SS are less related to each other than they are to BSS (and its close relatives), suggesting that the original XSS diverged in different directions to accommodate new substrates.

Enzymes within the HPD cluster may also accept multiple substrates given that, like the XSS enzymes, the nodes within this cluster are arranged into at least two distinct groups. There is some speculation (5, 6) that in addition to *p*-hydroxyphenylacetate conversion to *p*-cresol by HPD, a GRE may catalyze decarboxylation of tryptophan-derived indole acetate to form skatole

(3-methylindole), a known amino acid degradation product of various fecal bacteria that can induce pulmonary disease (7). This reaction, and other similar amino acid decarboxylations, may explain some of the diversity within this subgroup.

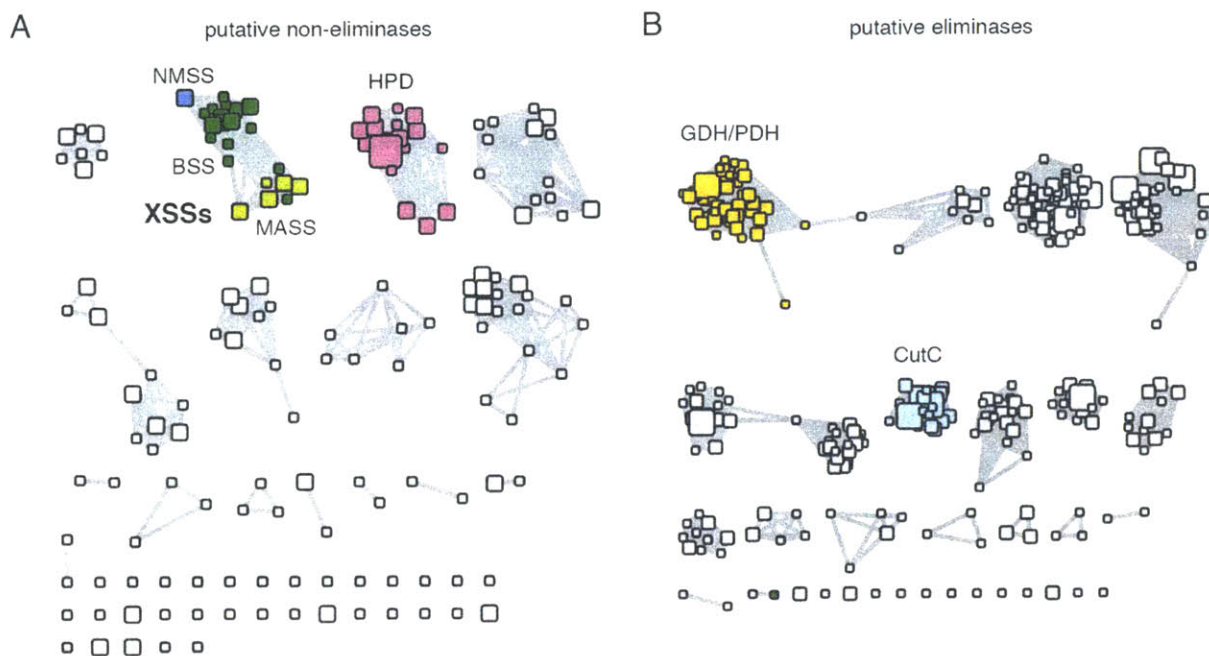


Figure VI.2. Sequence similarity network of functionally distinct subsets of GREs. (A) GREs not within the putative eliminate or PFL groupings (all gray nodes in Figure VI.1). Nodes were clustered at 80% sequence identity with edges drawn at ~30% identity. Labeling/coloring of the XSS cluster is according to database annotations of sequences from characterized bacterial isolates. **(B)** GREs denoted as putative eliminases (yellow nodes in Figure VI.1) based on the reactions performed by three characterized members. Nodes were clustered at 80% identity as in **(A)** but the edge threshold was raised to ~42% identity. At 30% identity, all nodes are connected. In both **(A)** and **(B)**, small nodes contain 1-2 sequences, medium nodes contain 2-20 sequences, and large nodes have over 100 (max ~3000). Abbreviations are as in Figure VI.1; NMSS: naphthyl-2-methylsuccinate synthase, MASS: 1-methylalkylsuccinate synthase, PDH: 1,2-propanediol dehydratase.

The other node clusters within Figure VI.2A are likely to represent enzymes with functions distinct from previously characterized non-eliminases. These enzymes have diverse Cys loop sequences, which suggests they may catalyze a wide range of chemistry as the Cys loop commonly contains motifs specific to enzymes that perform a particular type of reaction. Even within these clusters, sequence identity between members can be 25-35%, suggesting that not all clusters will be monofunctional. Indeed, the BSS-containing cluster includes enzymes that all

generate succinate-conjugated hydrocarbons, but the substrate range of these enzymes is impressive, and continues to expand with further characterization.

In the original sequence similarity network (Figure VI.1), a subset of GREs, including CutC and GDH, form a large group of nodes that are typically 35-45% identical. I have provisionally labeled these enzymes **putative eliminases**. Increasing the edge threshold to ~42% identity reveals nine large sequence clusters with at least ten nodes (Figure VI.2B). To illustrate the breadth of sequences and speculate on some of the associated diverged functions, I have constructed multiple sequence alignments with representative members of each cluster of the eliminase subgroup (Figure VI.3). These sequences contain the active site consensus motif GChE (where h is hydrophobic, typically V or I), consistent with a designation as dehydratase enzymes (although this sequence does not necessarily confirm the assignment). In both CutC and GDH, the Cys loop glycine allows binding of a primary alcohol adjacent to the active site cysteine, and the glutamate facilitates dehydration by accepting a proton. Two clusters do contain alternate sequences (GCVV or GCIQ) suggesting that these enzymes either 1) catalyze dehydrations using an alternate proton acceptor, 2) perform a different type of reaction altogether, or 3) are non-functional enzymes. Within GRE eliminases, $\beta 1$ and $\beta 6$ are likely to be especially informative regions with regard to function and substrate specificity. The structure of CutC and subsequent comparisons with GDH has revealed that these two strands contribute residues that enable binding of and catalysis with different substrates, glycerol for GDH and choline for CutC (Chapter IV). Notably, $\beta 1$ contains many insertions and deletions that likely remodel the protein backbone, as observed in the structure of CutC, as well as exchanging residue side chains.

(Figure next page)

accession	β 1	Cys loop	β 6	Gly loop	
gi 316924546	DND IKIECAPGETT	W SCIGC VEPL	SQW- SSAGHY	LLVRV AY SAYF	
gi 589393934	-IDK KAVRGWGEVT	W NPGGC VE TN	RAY- SALADI	LLVRV AY TAFF	
gi 226903820	-IDR KAVRGYGE NT	W NPCGC VE TN	KQY- TDIADI	LLIRV AY TAFF	
gi 531207381	DND SKWRQAVGE VT	W SNCGC V VP H	GQW- TSAVNI	LLVRV AY STQF	
gi 588301000	D VTIKTQ NGGG DTC	Y CLMGC VE PQ	YQW- TSVGYT	LMIRV AY S AFF	CutC-group
gi 218372442	DLS YHQV NGGG DTC	Y CLMGC VE PQ	YQW- TSTGYT	LIVRV AY S AFF	CutC-group
gi 78358322	DCS YHAVN GGG DSN	Y CLMGC VE PQ	YQW- TSTGYT	L VVRV AY S AFF	CutC-group
gi 219868924	DCS YHALN GGG DSN	Y CLMGC VE PQ	YQW- TSTAYT	L VVRV AY S AFF	CutC-group
gi 308917255	HRGGISMSG DGH IV	Y CPIGC VE TG	GHRAT GMTYV	LIVRV AY CAQF	
gi 308750777	HRGGISMSG DGH II	Y ASMG CV ETA	GHRAT GMTYV	LMVRV AY CALF	
gi 257050627	HGEGNMTSG DGH IV	Y SMVG CV EVA	GYR CTG MT F L	IIVRV AY CAQF	
gi 319811879	GMEGKMNSG DAH LA	Y SAIG CV ETA	GYR CTG MS Y M	LIVRV AY S AFF	
gi 225860290	GMEGKMNSG DAH LA	Y SAIG CV ETA	GYR CTG MS Y M	LIVRV AY S AFF	
gi 635829391	KAEGNMTSG DAH LA	Y AAIG CI ETA	GYR CTG MS F I	L VVRV AY S AFF	
gi 686280454	KAEGNMTSG DAH LA	Y ASIG CI ETA	GYR CTG MS F I	L VVRV AY S AFF	
gi 507814537	VL G NR-TQPSSGVV	W TVVGC VE PH	DAS-PIAG Y I	L VVRV AY L AYF	
gi 547730842	TV GLR -NGVSG ETT	W TIVGC I Q PG	DGS-PDAG Y V	LMVRV AY MAYF	
gi 363620213	TIGSR-DSAT G HIT	W TIVGC VE PH	DGS-PTSG Y L	L VVRV AY M AYF	
gi 494119912	TE- FMEQR GP G HTA	G GTSG CV EAG	AY--ILT G YF	LIVRV AY SDYF	
gi 476635646	SV- AQ TDK GQGH II	Y AVVGC VE LS	YGL- HDI S M F	LIVRV AY S A MF	
gi 328455416	SI- AQ TDK GQGH II	Y GVVGC VE LS	YGL- HDI S M F	L VVRV AY S AMF	
gi 662575791	SL- NQ TDK GQGH II	Y AVVGC VE LS	YGL- HDI A M F	L VVRV AY S AMF	
gi 261746299	NL- NQ TDK GQGH II	Y SVVGC VE LS	YGL- HDI A M L	L VVRV AY S AFF	
gi 508207507	KL- NQ TDK GQGH II	Y ATVGC VE T S	YGL- HDI A L F	LLVRV AY S A FF	
gi 635771027	SV- NQ TDK GQGH II	Y AVVGC VE LS	YGL- HDI A M F	L VVRV AY S AFF	
gi 488773305	KL- NQ TDK GQGH II	Y SIVGC VE LS	YGL- HDI A L F	L VVRV AY S AFF	
gi 672403675	SL- NQ TD HAQGH IL	Y SVVGC VE LS	LGW-SDA S M F	L VVRV AY S AFF	
gi 497274155	KI- NQ KD HAQGH IC	Y AIVGC VE L T	LGW-SDA A M F	LTVRV AY TAYF	
gi 330836088	KI- NQ KD HSQGH IC	Y AIVGC VE L T	LGW-SDA A M F	LTIRV AY TAYF	
gi 480705495	DDT WV SAAGL G NV I	W VPDGC IE A S	DFA-AKG P W L	LMVRV S Y S ALF	
gi 302336767	DDT WCA AAGL G NV L	W APVGC IE A T	DFA-AKG S W L	LMVR I S S Y S AFF	
gi 406960839	DDT WV SSAAL G N H M	W APVGC IE A T	DYA-AKG P W L	LLVRV S Y S CLF	
gi 478485187	Q TNLY HYAGAG H LA	Y CVVGC VE I S	YGW-HDA A Y V	MLVRV AY S A YF	
gi 546486902	SE Q LYTYNGV G H V C	Y VVGC VE P S	YGW-CD S A Y F	LMVRV AY S A YF	
gi 480699086	N PLSR TRSG H Y L	Y AIVGC VE P T	MGS-T N A G F F	L VVRV AY S AFF	
gi 663077408	D TRSR ARS G Y G H Y I	W AIVGC VE P T	MGR-T N S C H F	L VVRV AY S AFF	
gi 289502943	L GGNY FF TG V G H V C	Y CIIG CV E PD	YGW-HD A M F F	L VVRV AY S V F F	GDH-group
gi 239906065	TV GNY F Y NGV G H I S	Y GII GC V E P Q	EGW-HD A A F F	L VVRV AY S AFF	GDH-group
gi 658089048	TV VANY H F NG L G H I S	Y AIVGC VE P Q	DGW-YD A A F F	LIVRV AY S A QF	GDH-group
gi 324031271	TP GNY F Y NGV G H V T	Y NIIG CV E P Q	EGW-HD A A F F	L VVRV AY S ALF	GDH-group
gi 401705407	TV GNY FF G V G H V S	Y CIIG CV E P Q	EGW-HD A A F F	L VVRV AY S AQF	GDH-group
gi 315621024	TV GNY Y F GG I G H V C	W LPIG CV E P Q	DGW-HD A A F F	LIVRV AY S A QF	GDH-group
		:		: :*: *	

Figure VI.3. Sequence alignment of putative eliminase regions predicted to impact substrate specificity and catalysis. Sequences were chosen to represent a diverse group of nodes within the clusters of nodes denoted as putative eliminases (Figure VI.2B). Sequence positions that contribute specificity-determining side chains to the active site (bold) have been inferred from the structures of CutC and GDH. It should be noted that although most of these positions align within the sequences, the residues do not necessarily adopt the same three-dimensional structure. The alignment was generated by Clustal Omega (8). The Gly loop is included to illustrate the high sequence conservation in this region of the enzyme. The cysteine and glycine residues that form radical species during catalysis (red highlight) are absolutely conserved, along with several other residues in the Cys and Gly loops. Exceptions to conservation discussed in the text are highlighted (yellow).

The overall sequence identity of the complete sequences from the chosen putative eliminases ranges from ~29% to ~90%, but the median identity is 35%. This distribution of similarity is easily visualized as a heat map of the percent identity matrix generated by pairwise comparison of selected sequences (Figure VI.4A). Broadly, this matrix shows boxes of more conserved sequences (lighter colors) along the diagonal, as sequences are clustered according to similarity. Off the diagonal, there are sporadic associations between these boxes of 40-50% identity; however, most sequences are 25-35% identical overall. Importantly, although the sequences were chosen to be representative of the eliminase clusters, with large clusters contributing numerous sequences and smaller clusters contributing just a few sequences, the alignment is independent of the prior clustering of the sequences.

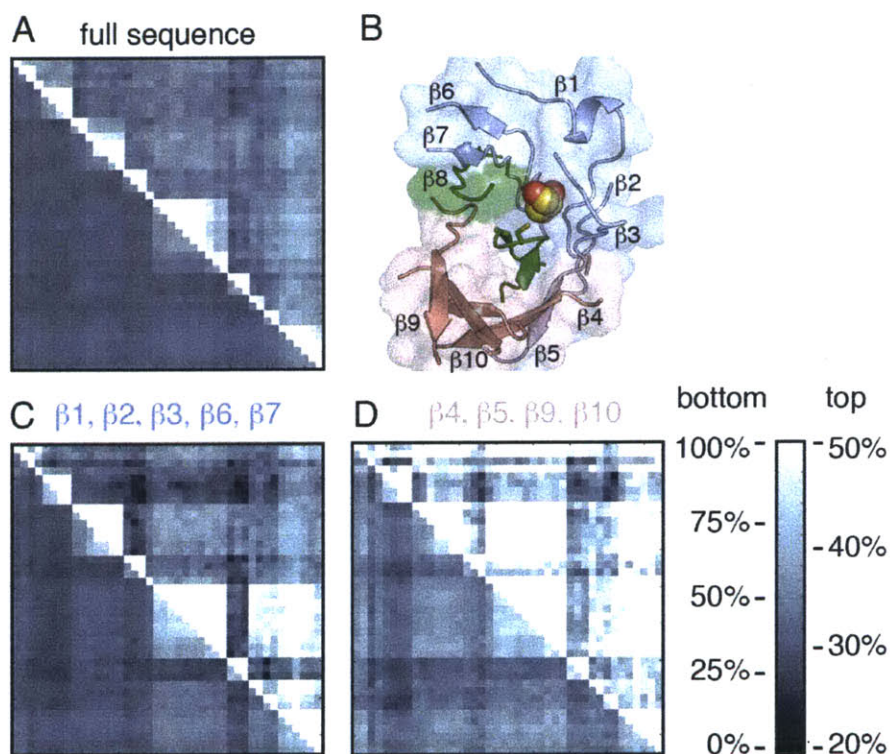


Figure VI.4. Sequence conservation within functional regions of putative eliminases. (A) Heat map of the percent identity matrix for the complete GRE sequence (~800 residues). (B) The structure of GDH with glycerol bound illustrates how residues on the upper portion of the β/α barrel (light blue) are involved in binding glycerol (spheres), although strands on the lower portion of the barrel (pink) contribute to the stability of the Cys loop (green ribbon) and its interactions with the Gly loop (not shown). $\beta 8$ (green) does not appear to be as involved in substrate binding in GDH. (C) Heat map of the percent identity matrix for strands on the top half of the GRE barrel that contribute to substrate binding: $\beta 1$, $\beta 2$, $\beta 3$, $\beta 6$, $\beta 7$, and the Cys loop

(~120 residues); **(D)** As in **A** and **C** but for strands on the bottom half of the GRE barrel that contribute to Gly loop and Cys loop interactions: $\beta 4$, $\beta 5$, $\beta 9$, and $\beta 10$ (34 residues). A color key with percent identity is given at the far right (same scale for all three panels). Percent identity matrices were generated by Clustal Omega (8) using the sequence accessions given in Figure VI.3.

When the Cys loop and upper strands of the barrel ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 6$, and $\beta 7$), which harbor residues important for substrate binding and catalysis (Figure VI.4B), are considered separately, however, the conservation within groups increases, whereas between groups it decreases. The median identity is slightly higher (37%), but it is apparent from the heat map (Figure VI.4A) that this increased similarity is mostly due to large increases in identity along the diagonal (closely related sequences), whereas off-diagonal identity is moderately increased in some regions and decreased in others (often 5-10 percentage points). Two large regions (bottom right) have increased similarity within the upper strands relative to the overall sequence, suggesting that they may operate on similar substrates despite a divergence in the overall sequence. Most off-diagonal sequences, however, see a drop in relative identity, indicating that these regions of the protein are more likely to diverge than the overall sequence.

When only the strands that lie on the bottom side of the barrel (Figure VI.B) are considered ($\beta 4$, $\beta 5$, $\beta 9$, and $\beta 10$), the median identity rises to 53%, and inspection of the heat map reveals an across-the-board increase in identity (Figure VI.4D). There are only a few sequences that appear to decrease in conservation relative to the complete sequence. This pattern of sequence conservation and divergence supports the conclusion that different regions of the GRE barrel experience different selective pressures. The upper strands are more highly conserved relative to the complete sequence within groups that appear to be isofunctional, but are less conserved between most groups, suggesting that the founding members of groups have diverged over time to accommodate new function. In contrast, the lower strands, which do not contribute to substrate binding, but instead form a stable platform on which the Gly and Cys loops rest, show less variability across the entire family. This pattern of conservation is in agreement with the idea that the initial hydrogen-atom abstraction between the Gly loop and Cys loop is likely to be similar, even in GREs that catalyze totally different transformations. The protein sequence corresponding to the structural elements that perform this step is therefore likely to be under greater purifying selection, preventing divergence within the Gly loop and associated regions of the enzyme.

Discerning the function of the sequence clusters defined in Figure VI.2B from their primary structure alone is a challenging task. Substantial backbone rearrangements due to insertions and lack of conservation within the central strands appear to be common in GREs, as discussed for the eliminases CutC and GDH in Chapter IV. Therefore, attempts to construct homology models that are informative are likely to be of limited value except for very close homologs as these models typically do not predict changes in the backbone conformation well. An alternate route to discerning function that has been successful in many other enzyme families is to look for operon structures containing conserved genes associated with the metabolic function of the GRE. An example of the utility of this strategy within the GRE family is the well-documented *dha/grp* operon (9), which contains aldehyde and alcohol dehydrogenases used to process the product of GDH, 3-hydroxypropanal. The presence of a complement of microcompartment proteins known to be involved in sequestering the reactive aldehydes produced by dehydration of glycerol and 1,2-propanediol by AdoCbl enzymes aided in the discovery and identification of these genes as glycyl radical eliminases with function analogous to the AdoCbl enzymes. The discoverers of the *cut* operon, which encodes CutC, used similar logic to identify such a gene cluster within an organism known for anaerobic choline utilization (4). Although this bioinformatics strategy may be useful in future attempts to determine sequence-structure-function relationships, it must be stressed that there is no substitute for experimental validation of the function of these clusters.

VI.C The phylogenic distribution of GREs and GRE metabolic steps.

One additional perspective on GRE function comes from prior knowledge of the metabolism of bacteria phyla and species. I have regenerated the sequence similarity networks described above to highlight several major bacterial Classes that are well represented in the network. Within putative non-eliminases (Figure VI.5A), the classes δ -Proteobacteria and Clostridia dominate, but β -Proteobacteria, Bacteroidia, and Actinobacteria are also abundant within specific clusters. The XSS cluster contains a diverse group of bacteria, consistent with the fact that many bacterial groups are known to degrade hydrocarbons using BSS-like enzymes. Most clusters are dominated by one or two types of organisms, and there are also no nodes with a very large number of sequences; there is only a single node (in the HPD group) with more than 20 sequences. One possible explanation for this lack of redundancy is that these enzymes do not

catalyze reactions common to diverse metabolic types and are therefore restricted to species performing degradation of specific and poorly distributed compounds. This is certainly the case with XSSs, which initiate a specialized branch of metabolism, and may also be the case with HPD, which is thought to contribute to virulence of *Clostridium difficile* and related pathogens through the actions of the product, *p*-cresol (10). HPD and three other clusters are annotated as originating from a gastrointestinal or other human source, but four of the remaining uncharacterized clusters have no such association and therefore may represent specialized metabolism in the environment. It will be particularly interesting to study these enzymes and the bacteria that produce them in light of the fact that XSSs are major contributors to anaerobic bioremediation.

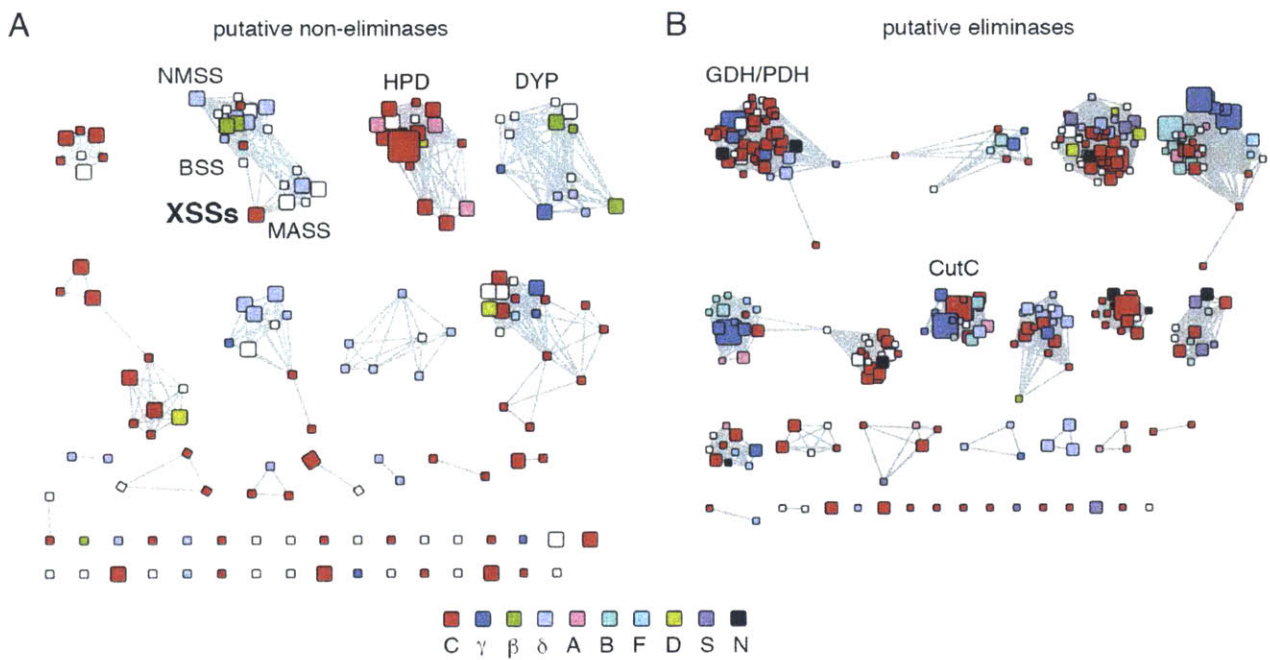


Figure VI.5. Phylogenetic breakdown of non-PFL GRE clusters. (A) Putative non-eliminases and (B) putative eliminases are shown as in Figure VI.2 but colored by bacterial Class as shown in the key. Major groups are C: Clostridia, γ : γ -Proteobacteria, β : β -Proteobacteria, δ : δ -Proteobacteria, A: Actinobacteria, B: Bacilli, F: Fusobacteria, D: Bacteroidia, S: Synergistia, N: Negativicutes. Other classes are in gray.

A second factor likely to contribute to the low apparent representation of putative non-eliminases in the sample of GREs analyzed here is bias in genomic databases toward microbes isolated from human samples. Better sampling of anaerobic environments, especially those that

present new metabolic opportunities to bacteria, may reveal new enzymes and new metabolic strategies.

Surprisingly, one of the non-eliminase clusters (DYP) contains genes from several aerobic fungal species and the Florida lancet, a vertebrate. These genes were almost certainly acquired by horizontal gene transfer as all of the other sequences are from aerobic, ocean-dwelling β -, γ -, and δ -Proteobacteria. These enzymes contain a heme-peroxidase domain at the N terminus and seem to be associated with other heme and copper proteins, and, in one case, an adenosylcobalamin-dependent light sensor. These enzymes are in fact the most diverged from the original cluster of PFL-like enzymes (Figure VI.1, far left offshoot), suggesting they have undergone a large transformation. Although these sequences appear to have a Gly and Cys loop, they are not associated with an AdoMet radical activating enzyme. All of these features suggest these enzymes may not be GREs as we know them but have either been salvaged to perform some other, non-radical reaction; found a different radical cofactor—perhaps heme; or found alternative way to generate the glycy radical in the first place. The function of these enzymes is not clear from their genomic context, but it is exciting to speculate that they could be performing new chemistry, either during activation or during the GRE catalytic cycle.

Within the putative eliminase subgroup (Figure VI.5B), sequences from enteric Proteobacteria dominate, with the two largest nodes containing over half of the unique sequences. Lactobacilli and Streptococci are well represented, along with many enteric Clostridia. Less common bacterial taxa represented in GRE clusters include those from the classes Actinobacteria, Spirochaetia, Synergistia, Negativicutes, Fusobacteria, and Bacteroidia. There are seven large nodes (> 100 sequences), two of which contain over 2,500 sequences and could thus represent broadly important metabolic enzymes. Although there is some grouping based on phylogeny, there is also evidence of LGT, as many sequences are >70% identical between different phyla.

Some Clostridia and most enteric Proteobacteria ferment sugars to generate energy, so it is tempting to speculate that some of the overlap within these groups could be from enzymes participating in sugar degradation, perhaps unusual sugars. This idea has precedent, as degradation of fucose is known to produce 1,2-propanediol, which is the substrate for propanediol dehydratase (11). Inspection of the gene neighbors of some of the uncharacterized putative eliminases reveals a large number of putative sugar transport proteins and

phosphotransferases involved in glycolytic degradation of sugars. It therefore seems likely that there may be additional sugar molecules that are broken down by radical elimination steps.

Clostridia are also notorious for fermentation of amino acids by the Stickland reaction, producing a number of foul-smelling byproducts. Indeed, the substrate for HPD, 4-hydroxyphenylacetate, is a product of tyrosine degradation by this pathway, although the role of HPD is not thought to be metabolic in nature. Again, inspection of gene neighbors has revealed proteases and branched-chain amino acid transporters associated with some sequences, suggesting that GREs could be involved in amino acid degradation in ways that are not yet appreciated.

VI.D The special case of pyruvate formate-lyase

Until now, PFL has not been considered in this analysis as there is far less diversity within these sequences, and all annotated enzymes contain a functionally important CCVS motif within the Cys loop (1) and are thus presumed to perform the pyruvate cleavage reaction. However, it is known that at least a subset of PFL enzymes, with the protein name TdcE, also operate on threonine-derived 2-ketobutyrate to produce formate and propionyl-CoA (12-14). It is unclear the extent to which this reaction is important for fermenting bacteria. There are few annotated sequences for this enzyme, and one of those that has been identified in *E. coli* is 82% identical to the *E. coli* PFL and catalyzes pyruvate cleavage in addition to threonine cleavage. Outside of this alternate substrate, there are no other known reactions catalyzed by the subset of enzymes that have the domain architecture of PFL (see Section I.F). Although many sequences are annotated as PFL, PFL2, or formate acetyltransferase, only sequences with the CCVS motif are likely to be true PFLs, as the second cysteine in the Cys loop is thought to be essential to the mechanism of C–C bond cleavage.

PFL sequences cluster in large part according to known bacterial phylogeny (Figure VI.6), although there is some evidence for LGT between phyla. Proteobacteria generally cluster together, with γ -Proteobacteria forming a distinct group from α - and β -Proteobacteria. Intriguingly, no sequences from δ -Proteobacteria are observed. As many δ -Proteobacteria are acetate-oxidizing sulfate-reducers that are incapable of growth on sugars, it is possible that the ancestor of this group lost PFL and adapted to a new way of life. Alternately, as δ -Proteobacteria

are thought to have diverged prior to the other classes, acquisition of PFL by LGT, perhaps from Clostridia could have occurred prior to their divergence.

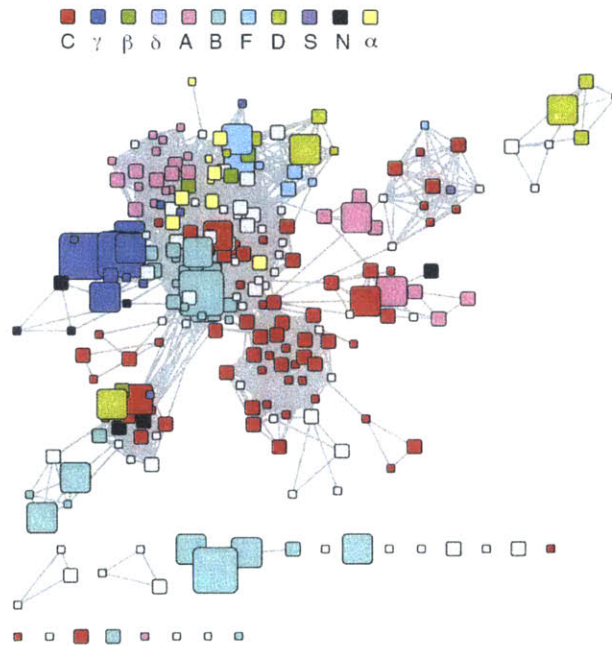


Figure VI.6. Phylogenetic distribution of PFL sequences. Nodes are clustered at ~85% identity and edges are drawn at 63% identity and above. Bacterial classes are colored as in Figure VI.5 with the addition of α : α -Proteobacteria

Within Gram-positive Firmicutes, there seems to be more diversity, with Clostridia and Negativicutes distributed in multiple groups, joined by sequence clusters dominated by Bacilli. The origin of this distribution is puzzling: it could be a relic of vertical decadence from before the Bacilli/Clostridia split or could represent LGT or gene displacement events. Interestingly, the largest γ -Proteobacterial cluster contains sequences from Bacilli, suggesting some extent of gene mixing between classes. Other groups show less divergence than these major phyla. Actinobacteria are split into two distinct groups, both closely related to nodes from various Clostridia. Bacteroidia are also split into several groups, not apparently closely related to either Bacilli or Proteobacteria, but instead are close to nodes from the more distant Fusobacteria and Spirochaetia. Taken as a whole, it appears that PFL is both much more slowly diverging than other GREs and less prone to LGT and gene duplication. The reasons for these differences are not at all obvious from this analysis and may be a fruitful area of future research.

VI.E The origin of GREs

Almost all the GRE clusters discussed above are exclusively bacterial. There are just six nodes with annotated archaeal origins, all within the dehydratase subgroup. No clear function has been assigned to these enzymes, although a structure is available for a putative eliminase from *Archaeoglobus fulgidus*. It thus seems likely that bacteria and archaea diverged prior to the diversification of GREs and that very few genes from this family were re-introduced to archaea by LGT. Likewise, only 16 eukaryotic nodes are present: 11 PFL nodes likely from two separate LGT events, and five highly similar nodes in the putative eliminase group, likely from a single, LGT event. There is evidence for functional eukaryotic PFL activity in *Chlamydomonas reinhardtii* (15) and anaerobic chytridiomycete fungi (16, 17) in association with hydrogenosomes, but the function of the eliminase-like enzymes is completely unknown.

As described above, more than half of GRE sequences cluster together as true-PFL enzymes (Figure VI.1, VI.6), despite there being much lower sequence divergence among these sequences. A large diversity of eubacteria is represented, even some sequences from organisms believed to be predominantly aerobic, such as cyanobacteria. The occurrence and high conservation of PFL in these organisms is puzzling, as there are numerous ways for pyruvate to progress in metabolism under both aerobic and anaerobic conditions. PFL plays a role in the central metabolism of saccharolytic, mixed-acid fermenters such as *E. coli* through conversion of pyruvate to acetyl-CoA and formate. However, many bacteria are not known to produce formate as an end product of fermentative metabolism (18). Instead, formate produced by PFL in these organisms could be an important anabolic metabolite that is incorporated into C1 metabolism through formyl-THF for use in nucleotide and amino acid metabolism (19-21). Recent studies have provided more evidence for the link between PFL and nucleotide biosynthesis, showing that PFL is upregulated by an increase in the putative global C1 sensor 5-aminoimidazole-4-carboxamide riboside mono- or triphosphate (ZMP/ZTP) (22). Some class III RNRs, particularly those from formate-producing, fermenting bacteria also use formate as a reductant for dNTP production from NTPs. In Chapter V, we discussed various reasons some bacteria would prefer an alternate source of reducing equivalents, such as NADPH, over formate for ribonucleotide reduction (18).

Many of the uncharacterized GREs described in Section VI.B and represented by smaller clusters are less abundant in genome databases, perhaps because they serve more specialized

roles and are thus less abundant in the environment or they are from habitats poorly represented by current sequence databases. Enteric bacteria are very well represented and also typically contain enzymes such as PFL and GDH, which are used in fermentative metabolism.

As discussed in Section VI.C, the occurrence of known GRE sequences (other than class III RNRs) in bacteria is predominantly within the phyla Proteobacteria (48%) and Firmicutes (37%). It is not known when GREs first evolved. Given their current distribution across many bacterial classes, they were likely present during the primary phase of bacterial diversification in the common ancestor of Proteobacteria and Firmicutes (23). The phylogenetic consistency of PFL sequences supports a primarily vertical descent of this gene, whereas other GREs appear to have been passed extensively among bacteria. GRE sequences from bacterial lineages believed to be early branching such as *Deinococcus/Thermus*, *Thermotoga*, or *Aquifex* are rare or nonexistent; however it is unclear if these species have lost these genes or they were never present. As GREs are inactivated in the presence of oxygen, bacterial lineages that have evolved to accommodate or require oxygen at some point in their evolution may lose ancestral GRE genes.

VI.F An explanation for the evolution of glycol radical enzymes

Within the GRE family there appears to be a huge range of diversity of sequence and function among non-PFL enzymes, even though true PFLs are very highly conserved. A simple explanation for this pattern of evolution is that PFL is an essential central metabolic enzyme and is thus under a great amount of purifying selection, preventing divergence. PFL is, however, not the only (or even necessarily the major) means of pyruvate utilization in bacteria: pyruvate dehydrogenase and pyruvate:ferredoxin oxidoreductase are equally or perhaps more important to energy conservation in most bacteria (24). However, PFL does play an important role in the production of formate for bacterial purine synthesis (20, 21), which could provide an alternate explanation for why PFL appears to be under such extreme selection relative to other GREs.

The second half of the explanation of why PFL and other GREs have such different evolutionary patterns is that GRE eliminases have apparently undergone extensive gene duplication and divergence. The genus *Escherichia* is represented in no less than eleven separate eliminase nodes in Figure VI.2B, including those for CutC (two nodes), and GDH (three nodes); the species *Clostridium difficile* is likewise represented in 14 different nodes in eight different

subgroups within the eliminase group. What is truly surprising is that, assuming an ancient origin for PFL, the highly divergent sequences present within the GDH cluster appear to have evolved solely from a single progenitor. Evidence for the single origin of non-PFL GREs comes from analysis of structures and sequences corresponding to the GRE dimer interface. In PFL, a distinct interface is formed from that observed in the non-PFL enzymes BSS, HPD, and eliminases. Were these enzymes to have arisen separately from PFL, they would not all share the common interface.

VI.G The distribution and evolutionary history of activating enzymes

Bioinformatic analyses (25) and structures of AdoMet radical enzymes have revealed a huge diversity not only in the core structure of these enzymes (26, 27), but also in accessory domains and cofactors (28) that enable expanded functions. Structural information for GRE activating enzymes (AEs) is currently quite limited, only one AE, PFL-AE, has been structurally characterized with its [4Fe-4S] cluster intact (29). PFL-AE is therefore the prototype for understanding how AEs bind and cleave AdoMet and interact with their GRE partners. However, the AE family is quite diverse in terms of both sequence and overall architecture (see Section I.G). Some AEs contain extra features that may impact their function or interactions with their substrate. In order to contribute an additional piece to the puzzle of GRE diversity, I examined sequences of a wide range of AEs. AEs are more conserved overall than their GRE counterparts, with up to 45% identity observed between PFL and non-PFL GREs (vs. ~20% identity between the corresponding GREs). Higher conservation within AEs is expected given that all AEs catalyze the same reaction and operate on the glycyl radical domain of their partner GRE, which is one of the most conserved regions of the protein between GRE families.

At first glance, the largest split within the AE family appears to be between the class III RNR-AEs (NrdGs), which are generally predicted to contain a smaller, five-stranded β/α barrel (Figure VI.7B), and the remainder of the family, all of which are thought to have six β/α units. However, a more detailed analysis based on a sequence similarity network (Figure VI.7A; constructed as described above) reveals that, in fact, the RNR-AEs are split into two groups that are apparently less related to each other than to PFL-AEs. In Chapter V (18), class III RNR sequences were divided into three groups based on the presence or absence a methionine, cysteine, and glutamate residue, each of which are proposed to contribute to catalysis. NrdD1

enzymes contain a MGCR motif within the Cys loop and are proposed to use formate as a reductant. NrdD2 enzymes contain a MCCR or ECCR motif as well as an EXTP motif that are thought to facilitate formate-independent reduction through disulfide formation in a manner analogous to that seen in class I and II RNRs. Finally, NrdD3 enzymes contain a QCCA or QCCR motif within the Cys loop, but do not have the EXTP motif. These distinct subclasses were correlated with the role of formate in metabolism rather than with phylogeny. Curiously, the sequence similarity network of RNR-AEs shows a slightly different pattern: AEs associated with bacterial, formate-utilizing RNRs form a distinct group (RNR-AE1) along with a large number of sequences not associated with any NrdD. The function of these AE-like enzymes is unclear, but several are from organisms that lack a NrdD and are known to lead a primarily aerobic lifestyle (e.g. cyanobacteria), so they may have developed an alternate function. Very few archaeal sequences are present in this group, and none are associated with a NrdD, suggesting this entire group may be of bacterial origin. If so, this could provide further evidence that the disulfide utilizing NrdD2 enzyme was the original class III RNR, as it is present in a wide range of bacteria and archaea.

A second group (RNR-AE2) comprises AEs for bacterial NrdD2s as well as all archaeal sequences, including NrdD1, NrdD2, and NrdD3 members. The analysis presented in Chapter V was somewhat ambiguous as to the relationship between archaeal NrdD; the corresponding NrdGs appear to be unambiguously related, although it is possible—albeit unlikely—that an early gene displacement has scrambled the relationship. Finally, an additional group of highly divergent sequences corresponds to AEs for bacterial disulfide utilizing RNRs, including those from *Thermotoga maritima* and *Neisseria bacilliformis*, described in Chapter V (18). These cannot be clearly placed in relation to the other groups in this network, but examination of the sequences suggests they are a split off of the RNR-AE2 group.

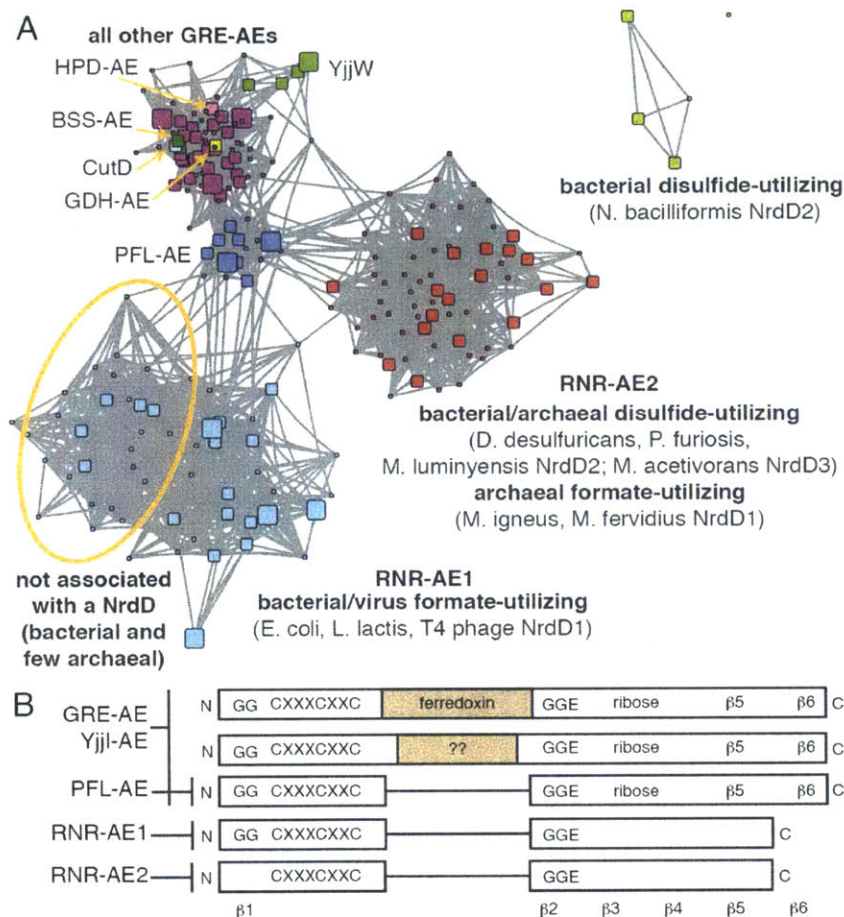


Figure VI.7. Sequence analysis of GRE activating enzymes. (A) A sequence similarity network of PFL- and RNR-AE sequences for class III RNRs (Interpro IDs IPR012839 and IPR012837) and all other GRE-AEs (Interpro ID IPR012839) displaying the diversity of and relationship between AE enzymes. Nodes (squares) represents a group of sequences clustered at ~40% identity. Edges connect nodes with at least 30% identity and 100 residue alignment length (typical AE length 150-200 residues). Distinct groups are apparent as described in the text: bacterial/virus NrdD1 RNR-AEs (teal), some of which are not clearly associated with a NrdD; archaeal NrdD1, NrdD2, and NrdD3 RNR-AEs (red); bacterial NrdD2 RNR-AEs; PFL-AEs (blue); Yjjl-AEs (YjjW, green); and other GRE-AEs (purple) that include AEs for BSS (dark green), GDH (yellow), CutC (light blue), and HPD (magenta). An additional group of bacterial NrdD2s is likely a branch of the RNR-AE2 group (yellow). **(B)** Some GRE-AEs have a domain architecture that includes a 2x[4Fe-4S] ferredoxin domain inserted after the conserved, [4Fe-4S]-cluster-coordinating CXXXCXXC motif. Some of these domains are missing cysteine residues that coordinate one or more of the [4Fe-4S] clusters. All PFL-AEs lack the ferredoxin domain but retain β6. Within β1, RNR-AE2s lack a GXG motif found in all the other AEs, suggesting that RNR-AE1 may be more closely related to PFL-AE. Both RNR-AE groups are highly divergent between β3 and the C terminus, although both groups appear to be missing large regions found in PFL-AE, and no clear motifs are found that could be responsible for AdoMet coordination in either subgroup. Additional structural information is needed to assign function to these regions.

The non-RNR-AEs are split into three separate groups: 1) the true PFL-AEs, which are associated with PFLs; 2) GRE-AEs, which activate a diverse range of GREs described in this chapter; and 3) YjjW enzymes that are associated with an uncharacterized enzyme that has some sequence similarity with GREs, Yjji (12). Yjji was not discussed above as it has low sequence identity with other GREs and has not been structurally or biochemically characterized, but it is very much worth investigation as it may represent new chemistry and/or a new domain architecture that sheds light on GRE evolution. Some YjjW and GRE-AE enzymes contain a ferredoxin-like domain (Figure VI.7B), whereas many do not contain this extra domain, suggesting a complicated origin or loss of the ferredoxin motif within the PFL-AE-like enzymes. Furthermore, whereas many of these motifs resemble the canonical ferredoxin motif, some have lost the cysteine residues required to bind one of the two clusters. Thus, this domain may have been gained early on in the evolution of GRE-AEs and subsequently lost if it did not confer an advantage.

The AE sequences provide an interesting link between the largest split in the GRE family: that between the PFL-like enzymes and class III RNRs. Because AEs are under different selective pressures—namely to maintain associations with AdoMet and their cognate GREs—they may offer a different perspective on GRE evolution. Structures of AEs with different architectures may be helpful in determining relationships between different groups and defining the minimal core fold of the AE family of AdoMet radical enzymes.

VI.H Conclusion

In this Chapter, my goal was to clearly define the diversity of GRE sequences now available through large genome databases. There is much room for future developments within this field. There are many enzymes that have yet to be characterized, most of which will likely catalyze new transformations that may be important to microbial metabolism. There may also be new domain architectures, oligomeric states, or protein-protein associations that provide additional functions to GREs, enable them to function in the presence of oxygen, or provide a new mechanism of activation. These speculations await biochemical characterization, but genetic and microbiological characterization will also be important to unraveling the true range of functions provided by these amazing enzymes. Finally, GREs are an ancient enzyme class, with likely roots in a pre-DNA world in the form of the class III RNRs. Understanding the evolution

of function within this family is a very difficult problem, but may shed light on microbial anaerobic metabolism as it has evolved over the past several billion years and continues to evolve and adapt to new environments.

VI.I Acknowledgements

Yifeng Wei and Smaranda Bodea have provided many insights and discussions about glycyl radical enzymes. Yifeng Wei has provided numerous ideas about microbial metabolism and its evolution. Prof. Daniel P. Dowling and Dr. Jennifer Bridwell-Rabb have helped in understanding the structure and evolution of AdoMet radical enzymes.

VI.J References

1. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.
2. Mitchell A, *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 43(Database issue):D213-221.
3. Atkinson HJ, Morris JH, Ferrin TE, & Babbitt PC (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4(2):e4345.
4. Shannon P, *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.
5. Yu L, Blaser M, Andrei PI, Pierik AJ, & Selmer T (2006) 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* 45(31):9584-9592.
6. Selmer T & Andrei PI (2001) *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of *p*-cresol. *Eur J Biochem* 268(5):1363-1372.
7. Carlson JR, Yokoyama MT, & Dickinson EO (1972) Induction of pulmonary edema and emphysema in cattle and goats with 3-methylindole. *Science* 176(4032):298-299.
8. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

9. Jorda J, Lopez D, Wheatley NM, & Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179-195.
10. Dawson LF, Stabler RA, & Wren BW (2008) Assessing the role of *p*-cresol tolerance in *Clostridium difficile*. *J Med Microbiol* 57(Pt 6):745-749.
11. Jorda J, Liu Y, Bobik TA, & Yeates TO (2015) Exploring bacterial organelle interactomes: a model of the protein-protein interaction network in the pdu microcompartment. *PLoS Comput Biol* 11(2):e1004067.
12. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.
13. Sawers G, Hesslinger C, Muller N, & Kaiser M (1998) The glycyl radical enzyme TdcE can replace pyruvate formate-lyase in glucose fermentation. *J Bacteriol* 180(14):3509-3516.
14. Hesslinger C, Fairhurst SA, & Sawers G (1998) Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol Microbiol* 27(2):477-492.
15. Atteia A, *et al.* (2006) Pyruvate formate-lyase and a novel route of eukaryotic ATP synthesis in *Chlamydomonas* mitochondria. *J Biol Chem* 281(15):9909-9918.
16. Akhmanova A, *et al.* (1999) A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism. *Mol Microbiol* 32(5):1103-1114.
17. Boxma B, *et al.* (2004) The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. *Mol Microbiol* 51(5):1389-1399.
18. Wei Y, *et al.* (2014) The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proc Natl Acad Sci U S A* 111(36):E3756-3765.
19. Thauer RK, Kirchniawy FH, & Jungermann KA (1972) Properties and function of the pyruvate-formate-lyase reaction in clostridia. *Eur J Biochem* 27(2):282-290.
20. Leibig M, *et al.* (2011) Pyruvate formate lyase acts as a formate supplier for metabolic processes during anaerobiosis in *Staphylococcus aureus*. *J Bacteriol* 193(4):952-962.

21. Derzelle S, Bolotin A, Mistou MY, & Rul F (2005) Proteome analysis of *Streptococcus thermophilus* grown in milk reveals pyruvate formate-lyase as the major upregulated protein. *Appl Environ Microbiol* 71(12):8597-8605.
22. Kim PB, Nelson JW, & Breaker RR (2015) An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell* 57(2):317-328.
23. Sheridan PP, Freeman KH, & Brenchley JE (2003) Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiology Journal* 20(1):1-14.
24. Ragsdale SW (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem Rev* 103(6):2333-2346.
25. Grell TA, Goldman PJ, & Drennan CL (2015) SPASM and Twitch Domains in S-Adenosylmethionine (SAM) Radical Enzymes. *J Biol Chem* 290(7):3964-3971.
26. Dowling DP, *et al.* (2014) Radical SAM enzyme QueE defines a new minimal core fold and metal-dependent mechanism. *Nat Chem Biol* 10(2):106-112.
27. Goldman PJ, Grove TL, Booker SJ, & Drennan CL (2013) X-ray analysis of butirosin biosynthetic enzyme BtrN redefines structural motifs for AdoMet radical chemistry. *Proc Natl Acad Sci U S A* 110(40):15949-15954.
28. Goldman PJ, *et al.* (2013) X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. *Proc Natl Acad Sci U S A* 110(21):8519-8524.
29. Vey JL, *et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

Michael A. Funk

77 Massachusetts Avenue, 68-694, Cambridge, MA 02139

+1 617 800 5352; mafunk@mit.edu

EDUCATION

Vanderbilt University, Nashville, Tennessee, USA

B.S., Chemistry, Biology, *Summa cum laude*, May 2008

Massachusetts Institute of Technology, Boston, Massachusetts, USA

Ph.D. Candidate, Chemistry. Successfully defended: March 2015

Research Advisor: Prof. Catherine L. Drennan

Thesis title: Structural Studies of Radical Enzymes in Bacterial Central Metabolism

AWARDS

Amgen Scholars Travel Fellowship (2014)

Howard Hughes Medical Institute Teaching Assistantship in Chemical Sciences (2010)

MIT Chemistry/Biology Interface Training Program Fellowship (2009)

National Science Foundation Graduate Research Fellowship (2009)

Joel Tellinghuisen Award for Outstanding Undergraduate Research (2008)

Phi Beta Kappa, Alpha Chapter of Tennessee (2008)

Amgen Scholars Research Fellowship, California Institute of Technology (2007)

American Society for Biochemistry and Molecular Biology Oral Presentation Invitation and Travel Grant (2007)

RESEARCH

Graduate Research, Massachusetts Institute of Technology (2009 – present)

Advisor: Prof. Catherine L. Drennan

Post-Undergraduate Research, Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany (2008 – 2009)

Project: Nuclear magnetic resonance and small angle X-ray scattering of ubiquitin

Advisor: Prof. Christian Griesinger

Undergraduate Research, Vanderbilt University (2007 – 2008)

Project: X-ray crystallography of a putative nitrosynthase enzyme

Advisor: Prof. Brian O. Bachmann

Undergraduate Research, Vanderbilt University (2005 – 2008)

Project: X-ray crystallography of G protein mutants

Advisor: Prof. Tina M. Iverson

Collaborators: Prof. Heidi E. Hamm, Prof. David P. Siderovski (UNC)

PUBLICATIONS

Funk, M.A., Marsh, E.N.G., & Drennan, C.L. (2015). Substrate-bound structures of benzylsuccinate synthase reveal how toluene is activated in anaerobic hydrocarbon degradation. Submitted to *Proceedings of the National Academy of Sciences*: April 2015.

Wei, Y., **Funk, M.A.**, Rosado, L.A., Baek, J., Drennan, C.L., & Stubbe, J. (2014). The Class III Ribonucleotide Reductase from *Neisseria bacilliformis* Can Utilize Thioredoxin as a Reductant. *Proceedings of the National Academy of Sciences*, 111(36), E3756-3765.

Funk, M.A., Judd, E.T., Marsh, E.N.G., Elliott, S.J., & Drennan, C.L. (2014). Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proceedings of the National Academy of Sciences*, 111(28), 10161-10166.

Ban, D.*, **Funk, M.A.***, Gulich, R., Egger, D., Sabo, T. M., Walter, K. F., Fenwick, R.B., Giller, K., Pichiem, F., de Groot, B.L., Lange, O.F., Grubmüller, H., Salvatella, X., Wolf, M., Loidl, A., Kree, R., Becker, S., Lakomek, N.A., Lee, D., Lunkenheimer, P., & Griesinger, C. (2011). Kinetics of conformational sampling in ubiquitin. *Angewandte Chemie*, 123(48), 11639-11642. (* Co-authors)

Ando, N., Brignole, E.J., Zimanyi, C.M., **Funk, M.A.**, Yokoyama, K., Asturias, F.J., Stubbe, J., & Drennan, C.L. (2011). Structural interconversions modulate activity of *Escherichia coli* ribonucleotide reductase. *Proceedings of the National Academy of Sciences*, 108(52), 21046-21051.

Vey, J., Al-Mestarihi, A., Yunfeng, H., **Funk, M.A.**, Bachmann, B.O., Iverson, T.M. (2010) Structure and mechanism of ORF36, an amino sugar oxidizing enzyme in everninomicin biosynthesis, *Biochemistry* 49(43), 9306-9317.

Preininger, A.M.*, **Funk, M.A.***, Oldham, W.M., Meier, S.M., Johnston, C.A., Adhikary, S., Kimple, A.J., Siderovski, D.P., Hamm, H.E., and Iverson, T.M. (2009) Helix dipole movement and conformational variability contribute to allosteric GDP release in α_i subunits. *Biochemistry*, 48(12), 2630-2642. (* Co-authors)

PRESENTATIONS

Funk, M.A. & Drennan, C.L. (March 2015) Substrate recognition by the radical enzyme benzylsuccinate synthase. American Society for Biochemistry and Molecular Biology Meeting, Boston, MA. Poster.

Funk, M.A., Chen, Y.T., & Drennan, C.L. (July 2014) Allosteric regulation by dATP in *E. coli* class Ia ribonucleotide reductase requires a change in quaternary structure. Enzymes, Coenzymes, and Metabolic Pathways Gordon Research Conference, Waterville Valley, NH. Poster.

Funk, M.A. & Drennan, C.L. (April 2014) Structural insight into anaerobic toluene activation by benzylsuccinate synthase. Howard Hughes Medical Institute Scientific Meeting, Chevy Chase, MD. Poster.

Funk, M.A. & Drennan, C.L. (June 2013) Breaking allosteric inhibition in ribonucleotide reductase. Building 68 (MIT Biology) Retreat, Falmouth, MA. Platform.

Funk, M.A., E.N.G., Marsh, & Drennan, C.L. (June 2013) Structural insight into anaerobic toluene activation by benzylsuccinate synthase. Boston Bacterial Meeting, Cambridge, MA. Poster.

Funk, M.A. & Drennan, C.L. (July 2012) Structural insight into the mechanism and regulation of benzylsuccinate synthase. Enzymes, Coenzymes and Metabolic Pathways Gordon Research Conference, Waterville Valley, NH. Poster.

Funk, M.A. & Drennan, C.L. (March 2012) Structural insight into the mechanism and regulation of benzylsuccinate synthase. Chemistry Student Seminars, Massachusetts Institute of Technology, Cambridge, MA. Platform.

Funk, M.A., E.N.G., Marsh, & Drennan, C.L. (January 2011) Structural insight into anaerobic toluene activation by benzylsuccinate synthase. Bioinorganic Chemistry Gordon Research Symposium, Ventura, CA. Poster.

Funk, M.A., Sintchak, M.D., Ando, N., Stubbe, J., & Drennan, C.L. (August 2010) The structure of a cobalamin-bound class II ribonucleotide reductase. American Chemical Society National Meeting, Boston, MA. Poster.

Funk, M.A., Oldham, W.M., Preininger, A.M., Hamm, H.E., & Iverson, T.M. (April 2007) The crystal structure of the fast exchange mutant I56C/Q333C in $G\alpha_{i1}$ suggests a mechanism for receptor-mediated allosteric nucleotide exchange. American Society for Biochemistry and Molecular Biology National Meeting, Washington, D.C. Platform and Poster

TEACHING

Guest Lecturer, Wellesley College

Introduction to Crystallography, CHEM 335, Physical Chemistry, Prof. Mala Radhakrishnan. (2013 – 2014)

Teaching Assistant, Department of Chemistry, Massachusetts Institute of Technology

5.78, Biophysical Chemistry Techniques. Graduate course. (2012)

5.111, Principles of Chemical Science. Undergraduate course. (2010)

5.07, Biological Chemistry. Undergraduate course. (2010)

Teaching Assistant, Office of the Dean of the Commons, Vanderbilt University

Organic Chemistry. Undergraduate course. (2007 – 2008)

Tutor, Chemistry Department, Massachusetts Institute of Technology

5.07, Biological Chemistry and 5.111, Principles of Chemical Science (2010 – 2013)

Tutor, Athletic Department, Vanderbilt University

General Chemistry, Introduction to Biology, Organic Chemistry, and General Physics. Undergraduate courses. (2006 – 2007)

Laboratory Training and Mentoring, Massachusetts Institute of Technology

- Kenny Kang, First-year Graduate Student. Investigation of the active complex of *E. coli* ribonucleotide reductase. (2013 – present)
- Yang-Ting Chen, First-year Graduate Student. Breaking allosteric regulation in *E. coli* ribonucleotide reductase. (2013 – present)
- Martin McLaughlin, Undergraduate Student. Structures of the AdoMet radical enzyme lipoyl synthase. (2013 – present)
- Kateryna Kozyrytska, Undergraduate Student. Capturing a substrate-bound form of the non-heme iron halogenase SyrB2. (2011 – 2012)
- Jose Ramirez, Summer Undergraduate Student. Snapshots of disulfide exchange at the active site of *E. coli* ribonucleotide reductase. (2013)
- Yasmin Chau, Rotation Student. Crystallographic analysis of ribonucleotide reductase mutants with altered activity regulation. (2013)
- Victor Cruz, Rotation Student. Crystallographic analysis of ribonucleotide reductase mutants with altered activity regulation. (2013)

Laboratory Training and Mentoring, Vanderbilt University

- Tarjani Thaker, Rotation Student. Expression and purification of $G_{\alpha_{i1}}$ protein. (2008)
- Chris Browne, Rotation Student. Construction, expression, purification, and crystallization of a mutant $G_{\alpha_{i1}}$ protein. (2007)
- Suraj Adhikary, Rotation Student. Expression, purification, and crystallization of $G_{\alpha_{i1}}$. (2007)

SERVICE

Graduate Activities

- Course 5 Teaching Assistant Training (2010 – 2014)
- MIT Student Support Services Proctor (2012 – 2014)
- MIT Chemistry Graduate Student Council, member and treasurer (2010 – 2012)
- MIT ChemBio Interface literature discussion leader (2010 – 2011)

Undergraduate Activities

- Vanderbilt Undergraduate Research Journal (VURJ) managing editor (2007)
- VURJ Engineering and Natural Sciences editor (2007 – 2008)
- Vanderbilt Student Volunteers for Science (VSVS) team leader (2004 – 2008)
- VSVS board member, historian (2007 – 2008)