AN ABSTRACT OF THE DISSERTATION OF

Camden McCullough Driggers for the degree of Doctor of Philosophy in Biochemistry & Biophysics presented on July 11, 2014.
Title: Structural Investigations to Understand the Mechanisms of Two Proteins Involved in Sulfur Chemistry: SsuE and Cysteine Dioxygenase.

Abstract approved: _____
P. Andrew Karplus

Sulfur is one of the six elements required during the early stages of the evolution of life, and enzymes involved in sulfur transfer and oxidation are increasingly being recognized as potential drug targets for antimicrobials as well as for therapies for cancer, neurodegenerative and inflammatory diseases. Bacteria are able to carry out a much broader range of sulfur chemistry than mammals and are able to use sulfate as a sole sulfur source, meaning that they can synthesize cysteine from sulfate, whereas mammals must take in either homocysteine or cysteine in their diet. Some bacteria, such as *Escherichia coli*, are able to obtain sulfate from alkanesulfonates through the expression a sulfur starvation utilization system, which includes a two-component system composed of an NADPH-dependent FMN-reductase, SsuE, and a monooxygenase, SsuD. Once cysteine (Cys) is available to a cell, the careful regulation of its concentration is important because even though Cys is required for life, at high levels it is toxic, especially for mammals. In bacteria, metazoa and fungi the Cys levels are primarily regulated by cysteine dioxygenase (CDO). In mammals, CDO is known to form a Cys-Tyr crosslink that greatly increases its enzymatic activity.

Here, I report structural studies and evolutionary considerations aimed at elucidating the mechanisms of SsuE and CDO. With regard to the work on SsuE, I

report the first structures of any SsuE, and these crystal structures show that SsuE is structurally similar to closely related FMN-reductases. These structures revealed that SsuE forms a tetramer that is similar to related FMN-reductases, and an active site that is completed by the dimer interface. An evolutionarily conserved π-helix appears to link FMN binding with tetramer dissociation. Three different states of SsuE were captured at ~ 2.0 Å resolution: an apo enzyme, an FMN-bound enzyme and a $FMNH_2$-bound enzyme. Based on these results, a reinterpretation of previous kinetics data on SsuE led to a novel proposal for the SsuE mechanism that is similar to those of its homologs. Furthermore, a general catalytic cycle was defined for NADPH-dependent FMN-reductases from the flavodoxin-like superfamily that provides a framework for understanding how the mechanism of these enzymes might change depending on cellular conditions and interactions with partner proteins.

With regard to the work on CDO, over 30 structures have been solved of wild type and mutant rat CDOs. In one set of studies, 14 crystal structures between 1.25 and 2.15 Å resolution, including a room temperature structure, are used to clearly define the pH dependence of Cys-persulfenate complex formation in the crystal, and fortuitously also provide the first high resolution view of an unreacted Cys bound in the active site. The main conclusions from this work are that persulfenate formation is consistently seen at pH values between 5.5 and 7, that is it not an artifact of freezing or synchrotron radiation, and at pH≥8 the unliganded active site iron shifts from 4- to 5-coordinate. We are able to identify that important active site pKa values lie between 5.0-5.5 and 7.0-8.0.

In a second study, 17 CDO crystal structures in the presence of Cys and inhibitors, ranging from 1.25 to 1.65 Å resolution and mostly of site-directed mutants, are used to define the role of an active site Cys-Tyr crosslink in catalysis as well as to determine the mechanisms of CDO inhibition by homocysteine and azide. Main conclusions are that a chloride ion binds to the active site iron in both the C93A and Y157F variants, even though the active site iron is still in the ferrous form. Upon

exposure to Cys, the chloride is displaced but Cys does not bind in the same way as it does to wild-type CDO, with the Cys largely coordinating the iron only through its thiolate, and not through the α-amino group. Cys-persulfenate does not form in the C93A or Y157F active sites, indicating the crosslink is necessary for persulfenate formation in the crystalline enzyme. These results defines a key role for Tyr157 in Cys binding, through both positioning Cys in the active site and modulating the pKa of its α-amino group. The structures in the presence of homocysteine and azide revealed why the CDO-homocysteine complex is unproductive for catalysis, and and how azide binds to the wild-type CDO, associating with the iron and the hydroxyl of Tyr157 in the crosslinked enzyme.

Finally, in a third CDO study, I describe the crystal structures of two bacterial CDO homologs that were originally solved by structural genomics groups but had not been reported in the literature. These are of great interest as they represent "Gln-type" and "Arg-type" classes of bacterial CDO homologs defined by the residue aligning with Arg60 of rat CDO. For the "Arg-type" CDO from *Bacillus subtilis*, we reproduced the crystals and were able to obtain a Cys-bound complex at 2.30 Å resolution that shows that its mode of Cys binding is comparable to mammalian CDOs. Also, we obtained the original diffraction images and further refined the structure of the "Gln-type" CDO homolog from *Ralstonia eutropha* to 1.65 Å resolution, and discovered the active site contained an unexpected iron-bound dioxygen. From this structure, we identified a novel active site Arg that is evolutionarily conserved among "Gln-type" CDO homologs and is positioned in a way that allows us to conclude that these enzymes cannot bind Cys and thus are not authentic CDOs. This is consistent with the observation that the one "Gln-type" CDO homolog with characterized substrate specificity has been identified as a 3-mercaptopropionate dioxygenases.

Structural Investigations to Understand the Mechanisms of Two Proteins Involved in Sulfur Chemistry: SsuE and Cysteine Dioxygenase

by

Camden McCullough Driggers

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented July 11, 2014

Commencement June 2015

Doctor of Philosophy dissertation of Camden McCullough Driggers presented on July 11, 2014.

APPROVED:

Major Professor, representing Biochemistry & Biophysics

Chair of the Department of Biochemistry & Biophysics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Camden McCullough Driggers, Author

ACKNOWLEDGEMENTS

First and foremost I thank my advisor, Andy Karplus. I admire your drive to constantly improve yourself and those around you and thank you for always pushing me. Thank you for sharing in the excitement of the breakthroughs of my research and for emphasizing that the obstacles provide "learning opportunities". Thank you for your positive support, patience, encouragement and resilience to adversity. Thank you for giving me the freedom to explore my ideas and for always keeping your door open to your students so that I didn't pursue the worst of those ideas. Thank you for teaching me to more effectively communicate science, and for modeling and teaching everyone a logical first-principled approach to understanding the universe. I have grown both personally and scientifically because of you.

I thank Dale Tronrud for being my guide through the world of crystallography and through the lava fields of Oregon. I am honored to be a good friend of the "Sage of the Willamette Valley". Thank you for the countless discussions on big ideas and for taking such an interest in my science. Thank you for keeping all of our computers running and for your patience with me when I would break them. Acknowledging you on every one of my publications seems insufficient to reflect the level of support you have given me both scientifically and personally. Thank you.

I would also like to express my gratitude toward Gary Merrill for recruiting me to the department, and for stepping up to chair our department. I thank Dan Arp for the use of your laboratory equipment, and for training two great scientists, Rick Cooley and Luis Sayavedra-Soto who have supported my science in various ways. I thank Steve Giovannoni for teaching me evolution and phylogenetic analysis, which was key to elucidating the mechanism of SsuE and advancing our understanding of CDO. I thank Ethan Minot, for your work to advance biophysical techniques, for supporting me as my graduate representative and for your student Landon Prisbrey, who always brought an interesting perspective to our "Structure Club". I would also like to thank my collaborators, especially Martha Stipanuk, Rick Cooley, Ryan Mehl, Holly Ellis, Larry Hirschberger, Victor Hsu and Ben Tribelhorn for all of your work.

Of my many lab-mates over the years I would like to thank Andrea and Justin Hall and Rick (and Stacy) Cooley, for your efforts to maintain our family-like friendship across the country. Rick, thank you for your humble collaborative spirit, for being so easy to communicate with and for all your help with my projects. Andrea, your outstanding yet humble mentorship was critical to the success I have had as a graduate student and every time I introduce a new student to the ropes I strive to be as good of an instructor as you were to me. Justin, you are a true leader and thank you for working to bring such a great group of people together. Thank you also to my friend and lab mate Russell Carpenter for all his tasty homebrew. I would like to thank my current lab-mates Kelsey Kean and Arden Perkins for your friendship and for helping to make the lab an enjoyable place to be. I also welcome Andrew Brereton to the lab. I would like to thank the undergraduates who let me practice my mentoring: Callia Elkhal, Justin Beil and especially Steven Hartman. Steven, you expressed such a positive attitude even when our project was not going smoothly, thank you.

I would like to acknowledge the National Institute of Diabetes and Digestive and Kidney Diseases grant number Grant DK-056649, the National Institutes of Health grant number R01-GM083136 and the National Science Foundation grant number MCB-0545048. I also am grateful for financial support from the Department of Biochemistry and Biophysics.

Mom and Dad, thank you so much for all the personal support and for nurturing my scientific curiousity. Jonathan ("Brudder"), thank you for challenging me intellectually, for help with the mathematics of crystallography and for the life-long friendship.

Lastly and most importantly, I thank my partner Sarah. Thank you for charming me during my first year and for showing me a hundred reasons to love Oregon, the greatest of which is you. Thank you for waiting that extra year to get into the Pharm. D. program at Oregon State so we could live together. Thank you for supporting me these last five years. Writing this dissertation and also (arguably) keeping my sanity would not have been possible without you.

CONTRIBUTION OF AUTHORS

Holly R. Ellis and Paritosh V. Dayal purified soluble SsuE and provided the analytical ultracentrifugation described in Chapter 2 and HRE provided input to the writing of the text. Rick B. Cooley contributed to the experimental design and data collection in Chapter 3 and provided input to the text. Banumathi Sankaran contributed to data collection in Chapter 3. Lawrence L. Hirschberger and Martha H. Stipanuk provided wild-type soluble cysteine dioxygenase for Chapter 3 and 4 and MHS contributed to the writing of Chapter 3. Steven Hartman contributed to the data collection and analysis in Chapter 5 and provided input to the writing of the text. P. Andrew Karplus was involved in the design, analysis and writing of all experiments and chapters.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

LIST OF APPENDIX FIGURES

Figure Page

Dedicated to

My Sarah

Structural Investigations to Understand the Mechanisms of Two Proteins Involved in Sulfur Chemistry: SsuE and Cysteine Dioxygenase

Chapter 1

Thesis Overview

Protein Structure-Function Studies

Proteins are the most diverse of all macromolecules and are essential to all forms of life. Each cell contains thousands of different proteins, which are each fine-tuned by evolution to serve a particular function. Among their many vital roles, they provide structure (e.g. keratin), small molecule transport (e.g. myoglobin), defense against infection (e.g. antibodies), molecular motors (e.g. dynein), and enable a virus to infect the host (e.g. gp120). One of the most fundamental roles of proteins is their ability to act as enzymes, which catalyze nearly all of the chemical reactions that occur in an organism, sometimes with reactions being catalyzed to such an extent that they are limited only by the diffusion of the substrate into the active site (e.g. superoxide dismutase). It is well established that for all proteins, their structure dictates their function. Thus it is the goal of structural biology to understand protein structure-function relationships.

Since the first protein structure was determined (Myoglobin; Kendrew, Bodo et al. 1958) the number of protein structures determined has continued to grow exponentially and this year the Protein Data Bank, a repository for protein structures, reached the 100,000th structure benchmark, with five of the structures included in this dissertation (PDB entries 4PIX, 4PIY, 4PJY, 4PIZ and 4PIY) being among those in the weekly release which breached the mark. The most powerful and successful tool for determining the three dimensional structure of a protein is X-ray crystallography, and this year is the International Year of Crystallography for which the United Nations recognized "that humankind's understanding of the material nature of our world is grounded, in particular, in our knowledge of crystallography." The importance of crystallography is reflected in the fact that twenty-three Nobel Prizes have been awarded in the area, with the most recent being for structure-function studies of G-protein coupled receptors (2012, R. J. Lefkowitz and B. K. Kobilka) and the ribosome (2009, V. Ramakrishnan, T. A. Steitz, and A. E. Yonath).

Because of the power protein structures have on understanding human health, the NIH funded Protein Structure Initiative (PSI) invested heavily in structural

genomics research centers. These efforts were have accounted for more than 13,100 solved structures (PDB query June 2014). One consequence is that there now exist thousands of entries in the PDB for protein structures that have not been described in the peer-reviewed literature. As is, these structures are of limited value to the broader scientific community both because they will not show up in literature searches and because no expert having both knowledge of protein crystallography and the particular protein family has carefully vetted the structures for accuracy and for information content. As such, there exists now an opportunity for researchers with appropriate expertise to make more accessible the many structures from structural genomics, and this is the type of study that is presented in Chapter 5 of this thesis.

X-ray crystallographic studies are possible because proteins, under the right conditions, form crystals, and in fact it was experiments on crystals that first showed X-rays were waves and part of the electromagnetic spectrum (Max von Laue, 1914 Nobel Prize). The same forces that drive proteins to fold into structures (i.e. the hydrophobic effect and electrostatic and van der Waals interactions), assembles them in arrays of copies so ordered that the protein crystal can be described in its entirety with just one model of its asymmetric unit (ASU) and symmetry operators relating it to the other multiple billions ($\sim 10^{12}$ - 10^{15}) copies in a crystal. Indeed, it is this perfectly ordered array that gives rise to diffraction when shining X-rays on a protein crystal, with each reflection representing a lattice plane through the ordered array in the crystal. As the X-ray wave goes through the crystal, the intensity of the wave is amplified immensely where there is constructive interference, but is not detectible anywhere else. Although this gives measurable signal, the phase information of the X-ray wave is lost, and so, in the cases of the studies presented in this dissertation, it can be calculated from an existing similar protein structure (>25 % sequence identity). Because the intensities of the reflections can be calculated from a model, refinement works by changing the model until the calculated intensities from the model fit the data. As a cross-validation method, 5 to 10 % of the measured reflections are never included in refinement and are just used as a measure of the ability of the model to

predict the data. For small molecule crystallography, this is often not needed as the data are of such high resolution that the atoms can just be placed. However, protein crystal structures can still be determined that are quite precise. As example, one of the cysteine dioxygenase (CDO) structures described in Chapter 3 of this dissertation has coordinate uncertainties of ~0.05-0.10 Å for the C, N, and O atoms and even better for the sulfur and iron atoms (for scale a carbon-hydrogen bond length is ~1.1 Å).

In order to pack in the crystal, a protein makes intermolecular contacts that it does not make in dilute solutions, and in the early days of crystallography people wondered if the structure of a crystalline protein was the same as its structure when it was in solution. In fact, protein crystals are mostly solvent, with up to 80% of the protein's surface being exposed to solvent so that most of the surface is not influenced by the packing interactions. One concrete line of evidence that crystal packing interactions do not highly deform protein structures is that when the same protein packs differently in a crystal (i.e. a different crystal form) it has basically the same structure, and similarly, so do evolutionary-related proteins (homologs). Also, crystalline enzymes are also often active, and their structures agree well with inferences based on other methods of characterizing them. As NMR emerged as a technique for determining protein structures, it became clear that the crystal structures were typically equivalent, but more precise than the NMR derived models. Even so, the crystal packing interactions must be carefully examined during the course of the analysis of the structure, because they can have an impact on the structure that is not physiologically relevant. In the case of SsuE described in chapter 2, a crystal packing interaction occurs at the active site of two of the copies of the protein, making those particular active sites not representative of the in-solution behavior of the enzyme. However, in most cases the trained eye can determine if and how crystal packing interactions are impacting the structure.

Protein crystals are mostly solvent, this means that upon exposure to air they quickly dry out. Crystallographers can overcome this by keeping crystals hydrated in capillaries, or far more commonly they flash freeze the crystal at liquid nitrogen

temperatures (75 K) for data collection. This method of cryocrystallography has the added benefits of stabilizing the crystal so that it diffracts to higher resolution, and making the crystal more resistant to radiation damage. A decrease in temperature also decreases the thermal vibrations of the atoms of the crystallized protein, resulting in a more ordered crystalline structure. This is reflected in the crystallographic temperature factors (also B- or Debye-Waller factors), which represent the spread in the positions of the atoms caused by both thermal motion and the disorder of the crystal. Entropic contributions are also less at low temperature ($\Delta G=\Delta H-T\Delta S$), and because of the greater contribution of enthalpy, some dynamic effects (i.e. sidechain alternative conformations) are no longer seen (i.e. single conformations dominate and more waters can be seen) (1). Because of this a control structure at room temperature can be determined for comparison, as was done in Chapter 3.

Cryocrystallography is essential for data collection at synchrotron radiation sources, as it allows crystals to keep their internal structure and still diffract throughout the course of a data collection. Synchrotron sources are of particular benefit to the work here because they allow for higher resolution structures to be solved. The much greater flux of synchrotrons (2.5×10^{11} photons/(sec*mm^2)) compared to our laboratory x-ray generator ($\sim 10^3$ photons/(sec*mm^2)) allows for those weaker high-resolution (high angle) reflections to be measured. One possible problem that arises with this incredibly high flux is chemistry occurring to the protein or its ligands as a result of this radiation damage. The molecular effect of this radiation damage can actually be observed by collecting a highly-redundant data set on a crystal, then treating the first set of images and the last set of images separately, and calculating a Fobs(first)-Fobs(last) map. When this particular experiment is done for CDO crystals, decarboxylation of Glu and Asp sidechains can be seen, but no radiation induced chemical changes are seen in the active site. Another typical feature of radiation damage is the reduction of flavins, iron, and disulfide bonds. For this reason crystallographers often do lower energy data collections on laboratory based X-

ray generators to help to minimize radiation damage when it is of particular concern (Chapter 3).

The second reason that synchrotron sources allow higher resolution structures to be solved is that they have a shorter wavelength of X-rays. The wavelength of radiation is related to the resolution of the reflection by Braggs Law (wavelength = 2 x resolution x sin(scattering angle/2)), and so with shorter wavelengths, the scattering angle is smaller for a given resolution. Also, the highest possible resolution is half of that wavelength. Our laboratory X-ray generator has a Cu anode, which when struck by electrons generated by a cathode (tungsten filament), produces X-rays with a wavelength that is characteristic of the metal; for Cu this is 1.54 Å. The photon generating effect is called "Bremsstrahlung" radiation, and actually accounts for <1% of the energy of the electrons that strike the Cu anode, the rest of the energy goes to heat, so the Cu anode must be cooled in clever ways and much of the work in maintaining a laboratory based X-ray generator is in fixing those systems related to cooling the anode. Synchrotron sources generate X-rays very differently; they generate photons by accelerating electrons radially. Typical protein crystallography data collection at the synchrotron occurs at a wavelength of 1.00 Å (12.7 keV). In these studies, using cryo-crystallography and collecting at synchrotron sources was necessary to obtain high enough resolution to gain some of our insights, but also had to be controlled for to make sure that the Cys-persulfenate seen was not an artifact of these techniques. In the case of the *Bacillus subtilis* CDO structure described in chapter 5, lacking high resolution data made us unable to conclude what was bound in the active site.

Another recent advance in crystallography helped extend the resolution of the crystal structures reported in this dissertation. Recently, Karplus & Diederichs (2) introduced a data quality indicator termed $CC_{1/2}$ and showed by way of a paired refinement strategy that the ability of the model to accurately predict the data is improved by extending the high resolution limit beyond those indicated by the current conventional standards of the signal-to-noise ratio, $\langle I/\sigma \rangle$, being at least 2, and a data-

quality indicating R-factor, R_{meas}, being less than about 60%. Their main test case was one of the CDO data sets collected for this work (Chapter 3, the Cys soak at pH=6.2), and they showed that improved models were obtained using data out to $CC_{1/2}$ between 0.1 and 0.2 even though the merging R-factors become very high. In order to create the best models for the structures in this dissertation, we used such $CC_{1/2}$-based cutoffs throughout and found that it improved the structures in all cases.

Solving a structure allows for accurate phylogeny of very diverse proteins; this is because structure is conserved more than sequence. The natural process that drives evolution occurs at the level of DNA, with natural sources (i.e. electromagnetic radiation from the sun, oxidative damage) causing DNA damage and cells not perfectly repairing that damage, such that errors are constantly occurring. Many mutations do not change the protein at all, because of the redundancy of codons, meaning that there are more variations in the DNA than in the protein sequence. Also, among the amino acids there are substitutions that are conservative (e.g. Ile to Val or Leu) such that the structure of the protein does not change, meaning that there are more changes in the sequence than in the structure. In this way, the genes encoding two related proteins can be rather divergent on the DNA level, but the encoded proteins may have very similar structures and functions. Because protein structure is quite conserved, scientists have developed accurate programs for overlaying structurally similar proteins. One example is the Dali database (3). This tool is made available to the community so that a new structure can be compared against this database, and the potential phylogenetic relationships of the protein can be quickly elucidated.

In the studies reported in this dissertation, examining diverse groups of related proteins allowed us to see the variations in structure that occurred as a result of evolution, and allowed us to speculate about how, in one case, an enzyme should have a similar function to its relatives (SsuE, Chapter 2) and in another case that an enzyme had evolved a different function from its homologs (*Re*CDOhom, Chapter 5). Thinking with an evolutionary perspective at the structural level allows for such

powerful insights that have, as described in Chapter 2 and 5, allowed us to correct errors in the field and move our understanding of sulfur acquisition and metabolism forward.

Graduate work not included in this dissertation

In addition to my own structural biology work on the sulfur chemistry enzymes included in this dissertation, I was also an author on three peer reviewed publications for which another scientist was the lead author. I was also lead author on a short conference proceeding. I will outline these studies briefly below:

R.B. Cooley, J.L. Feldman, **C.M. Driggers**, T. Bundy, A.L. Stokes, P.A. Karplus, and R.A. Mehl. (2014) Second generation 3-nitro-tyrosine tRNA synthetases: the importance of selection strategy in non-canonical amino-acyl tRNA synthetase development. *Journal of the American Chemical Society* 53(12):1916-24

The Mehl lab had designed a 3-nitro-tyrosine tRNA synthetase (nitroTyr RS) that were an order of magnitude better at incorporating nitroTyr than previous nitroTyr RS. One aspect of this publication explored the structural basis for this improvement and found that the second generation nitroTyr RS active site is better suited to bind the substrate, but suggested the interaction of the nitroTyr RS with the tRNA also plays a role. For this study, I collected diffraction data and provided some input to the analysis of the structures and to the text.

C.M. Driggers, Ellis, H.R., & Karplus, P.A. (2012) Crystal Structure of Escherichia coli NADPH FMN reductase SsuE with and without bound FMN. *Proceedings of the 17th International Symposium of Flavins and Flavoproteins*. Miller, S., Hille, R., Palfey, B., Cecchini, G. Eds.; Alan R. Liss: New York.

This was a preliminary conference proceeding on the crystal structure of SsuE that we published before the refinements were completed and before the insights into mechanism were elucidated (as described in Chapter 2). This preliminary report

featured the alternative interpretation of the electron density maps of EmoB, SsuE's closest known structural homolog. Because I regularly attended conferences throughout graduate school, the EmoB authors knew me personally and were amenable to examine our alternative interpretation, which they agreed was correct. The original coordinates were obsoleted and the new alternative coordinates were deposited.

Berkholz, D. S. **C.M. Driggers**, Shapovalov, M. V., Dunbrack, R. L., Jr., and Karplus, P. A. (2012) Nonplanar peptide bonds in proteins are common and conserved but not biased toward active sites. *Proc Natl Acad Sci, USA* 109, 449-453

In this publication we showed that the peptide bond, which had been historically thought of as planar, actually deviates systematically up to 8 degrees from planarity, depending on the ϕ,ψ conformation. My part of this study was to verify the most extreme 116 examples of non-planarity by examining each model, analyzing them to see if they are biased towards being in active sites, which they are not, and to see if they are evolutionarily conserved, which they are. In addition to these analyses I provided figures and text for this publication. The ϕ,ψ conformational dependence was not enough to completely capture the degree of non-planarity observed, and we speculated that this could be due to a tertiary structure effect.

I also led a follow up study where we investigated one possible tertiary structure effect. The idea was based on a few examples where a repeating pattern of positive and negative omega angles along a twisted β-sheet was observed. We suspected that the twist of the beta sheet could be a tertiary structure effect contributing to non-planarity. In order to test this, I worked in collaboration with computer scientist Ben Tribelhorn (Assistant Professor, Seattle University) and after doing a survey of all high resolution proteins we found that there is not a correlation between degree of β-twist and non-planarity of the peptide bond except when using a manually selected subset as a control to test his program (data not published).

Fleishman, S. J., (62 others), **Driggers C.M.**, Hall R.G., Morgan J.L., Hsu V.L, (28 others), D. Baker. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *Journal of Molecular Biology*, 414, 289-302

The CAPRI experiment is a bi annual competition where I worked with Victor Hsu and two classmates using de novo computational methods to try and predict the hetero complex of two proteins. Each of the individual protein structures were known and the complex we were trying to predict had just been solved but had not yet been released. In this study, I wrote code using python and helped analyze some of the energy minimization data and provided input to the text. After the manuscript was submitted, an ongoing collaboration was forged with Ben Tribelhorn and Victor Hsu where I serve a minor advisory role, giving expert advice on protein structure.

In the remainder of this introductory chapter I will provide a brief introduction to general aspects of sulfur metabolism as it relates to SsuE and CDO. Then I will outline the contents of the remaining chapters of the dissertation.

Sulfur acquisition and metabolism

Sulfur is required for life as we know it, and enzymes involved in sulfur transfer and oxidation are increasingly being recognized as potential drug targets for antimicrobials as well as therapies for cancer, neurodegenerative and inflammatory diseases. (4-7)

No bacteria can grow without a sulfur source to be able to synthesize cysteine, but for aerobic bacteria like *Escherichia coli* and *Bacillus subtilis*, sulfur acquisition is particularly energetically costly (8) as sulfate must be actively transported into the cell against a negative intracellular potential (-70 mV), and then be converted from the highly oxidized sulfate to the highly reduced cysteine in the presence of oxygen. This is challenging for the bacteria and causes them some oxidative stress (9-10).

Sulfate assimilation begins with the active transport of sulfate across the bacterial cell membrane by an ABC-type sulfate transporter. Once inside the cell, ATP sulfurylase catalyses the reaction of sulfate with ATP to generate adenosine-5'-phosphosulfate (APS). Subsequently, APS reductase can then reduce APS to sulfite. Among many alternative pathways that exist, the shortest route to L-cysteine production is summarized in Figure 1.1.

Since sulfate is the fourth most abundant anion in human plasma (11), for bacteria which live inside the gut of a mammal they have plenty of sulfate, and APS levels are high (Figure 1.1). APS acts as a signal or sulfate levels, it binds as a negative cofactor to a LysR-type transcriptional regulatory protein called Cbl and inhibits its association with another protein CysB. However, for bacteria living in many other conditions, such as certain aerobic soil conditions, sulfate can be limiting. When sulfate is limiting then the concentration of APS in the cell is low, and the APS-free Cbl is able to form a heterocomplex with CysB, and this heterocomplex induces the translation of the sulfur starvation induced (ssi) genes (8) (reviewed in (12)), these ssi genes encode for the sulfur starvation utilization (Ssu) proteins.

Included on the cluster of ssiEABCD genes are genes that code for the SsuEABCD proteins. The proteins encoded by the A, B, and C genes make up an ABC type alkanesulfonate transporter, and those encoded by the E and D genes make a two-component monooxygenase system. In this two component system, SsuE uses NADPH to reduce FMN to $FMNH_2$ which it then transfers to its partner monooxygenase SsuD. SsuD is then able to use $FMNH_2$ and oxygen to cleave sulfite from the alkane sulfonate (Figure 1.2). A structure of SsuD is known and a proposed mechanism for SsuD chemistry proceeds through a peroxyflavin intermediate (Figure 1.3). This mechanism is similar to the SsuD homolog, luciferase, which is the well studied enzyme responsible for firefly luminescence.

SsuE has been studied enzymatically, but no structure was available before the studies reported in Chapter 2. SsuE was proposed to have an NADPH leading ordered sequential mechanism (13-14) (Figure 2.1B). Interestingly, this mechanism is different

from related enzymes from the flavodoxin-like superfamily (15), which have a ping-pong bi bi reaction mechanism (Figure 2.1C). This led us to wonder what structural features gave rise to this case of homologous enzymes having apparently divergent enzyme mechanisms.

Returning to the overall sulfur acquisition pathway shown in Figure 1.1, once sulfite is obtained, from either alkanesufonates or from sulfate, it can be further reduced by NirA (a sulfite reductase). When the structure of NirA was solved, it was realized that it forms an unusual Cys-Tyr crosslink in the active site that is essential for activity and is autocatalyzed when the enzyme first reacts with substrate (16). This kind of Cys-Tyr crosslink is found in only two other enzymes, one of which is cysteine dioxygenase and investigating the role of the crosslink in CDO is the focus of Chapter 4 of this dissertation.

After the sulfite is reduced to a sulfide (S^{2-}) by NirA, it can react with O-acetylserine in a reaction catalyzed by O-acetylserine sulfhydrylase to form cysteine. At last, the bacteria has successfully made cysteine can then be used for protein synthesis. Mammals lack the enzymes to convert sulfite to cysteine, and so they must either take in Cys directly or synthesize it from methionine.

Although cysteine is essential for life, it is also true that high intercellular levels cysteine are toxic because they generates reactive oxygen species (ROS) by driving the Fenton reaction (i.e. Fe(III)+ 2Cys$\rightarrow$ Fe(II) + cystine; Fe(II) + H_2O_2 $\rightarrow$ Fe(III) + $\bullet OH + OH^-$), and the ROS produced can cause toxicity through damaging proteins and DNA (9). In mammals, the cysteine toxicity is much more severe, because in addition to causing oxidative stress, Cys acts as an agonist of the NMDA-subtype of glutamate receptor, and cause an excitotoxic signal that leads to neuronal cell death, particularly in the CNS (17-20). High levels of cysteine and homocysteine have been observed in humans with Alzheimer's, Parkinson and motor neuron diseases (21). Therefore, even though the production of Cys is essential for life, its efficient removal is required to prevent cell death.

Bacteria, fungi and metazoa primarily regulate cysteine levels by the enzyme cysteine dioxygenase (CDO), which is an non-heme iron-dependent enzyme that in its ferrous form carries out the oxygen-dependent conversion of cysteine to cysteine sulfinic acid (CSA), and requires no external reductant for turnover (Figure 1.4). In mammals, CSA is committed to be further metabolized to taurine or to pyruvate and sulfite. In mammals, CDO is the most highly regulated enzyme in Cys catabolism, and in addition to the cell being able to control the amount of CDO enzyme (22), CDO also autocatalyzes a Cys-Tyr crosslink that greatly increases the catalytic efficiency of the enzyme. This crosslink formation is thought to represent a feed-forward activation in response to high Cys levels (23).

Bacteria are able to carry out a much broader range of sulfur chemistry on CSA than mammals (reviewed in (8)), however they can synthesize CSA other ways so until a study in 2006, it was not known if bacteria had CDOs (24). In that study, it was discovered that indeed bacteria have CDO enzymes, and that there exist a new kind of CDO homolog in bacteria, a "Gln type" CDO homolog that was not enzymatically characterized. These differed from the bacterial CDO enzymes that were characterized in the paper and also from the well characterized mammalian CDOs, which have an Arg at the Arg60 equivalent position and are called "Arg-type".

There are two other known types of thiol dioxygenases that are related to CDO, but have differing specificities. These are 3-mercaptopropionate dioxygenase (MPDO) and cystamine dioxygenase (ADO) (25). The two substrates involved are closely related to Cys, one being Cys without its α-amino group, and the other being Cys without its α-carboxylate. For rat CDO, 3-mercaptopropionate inhibits CDO activity by 50% at 1.2 mM, but cystamine actually enhances the activity, but does not act as a substrate (26), unless the active site Arg60 is mutated to an Ala (23). The three enzymes are all part of the cupin superfamily, and are fairly divergent homologs (<30% sequence identity).

This brief overview of the broader sulfur metabolism shows the relationship between SsuE sulfur acquisition and CDO cysteine catabolism. For the remainder of

the dissertation I will focus on SsuE and CDO. More detailed information on each of these enzymes is given in the introduction provided at the beginning of each main body chapter.

Dissertation Contents

There are five remaining chapters of this dissertation, including four chapters that present primary research projects, and a final chapter that highlights the impacts of this work and suggests future studies. In this section, I draw my introduction to the thesis to a close by providing for each of the four chapters in the body of the thesis information about where that work has been published or will be submitted for publication, as well as a brief summary of the work presented.

Chapter 2: "Crystal structure of Escherichia coli SsuE: Defining a General Catalytic Cycle for FMN reductases of the Flavodoxin-like Superfamily" Camden M. Driggers, Paritosh V. Dayal, Holly R. Ellis, and P. Andrew Karplus. Published in *Biochemistry*, 2014, 53 (21), pp 3509–3519

The major motivation for this structural investigation was to figure out the unusual difference in kinetic mechanisms reported for SsuE and its homolog EmoB. Three different structures of SsuE were solved: an apo enzyme, an FMN-bound enzyme and a $FMNH_2$-bound enzyme. The similar structures of SsuE and its homologs prompted a carefully re-evaluation of previous kinetics data, and it was discovered that the interpretation of the results from one key experiment were wrong. A reinterpretation of the results supported the conclusion that the kinetics data are actually consistent with SsuE having a mechanism similar to its close homologs. A general catalytic cycle was defined for NADPH-dependent FMN-reductases from the flavodoxin-like superfamily that provides a framework for understanding how the mechanism of these enzymes might change depending on cellular conditions and interactions with partner proteins.

Chapter 3: "Cysteine dioxygenase structures from pH 4 to 9: Consistent Cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH" Camden M. Driggers, Richard B. Cooley, Banumathi Sankaran, Lawrence L. Hirschberger, Martha H. Stipanuk, and P. Andrew Karplus. Published in *Journal of Molecular Biology* (2013), 425(17), 3121-3136

This study was the logical follow-up to a study published in 2008 (27), in which a never before seen Cys-persulfenate intermediate was captured in a single crystal. Because later quantum mechanical calculations led to suggestions that the Cys-persulfenate was not a true intermediate, it was important to learn more about the origins of its occurrence and stabilization in the active site of CDO crystals. This study reports 14 structures between 1.25 and 2.15 Å resolution, including a room temperature structure, that clearly defines the pH dependence of Cys-persulfenate complex formation in the crystal, and fortuitously also provides the first high resolution view of an unreacted Cys bound in the active site. The main conclusions from this work are that persulfenate formation is consistently seen at pH values between 5.5 and 7.0, and that is it not an artifact of freezing or synchrotron radiation. Also important was a conclusion that a previously published structure of human CDO with cysteine was actually not a reliable result.

Chapter 4: "Crystal Structures of Non-Crosslinked Variants of Mammalian Cysteine Dioxygenase With and Without Ligands" Camden M. Driggers, Lawrence L. Hirschberger, Martha H. Stipanuk, and P. Andrew Karplus. In preparation for *Biochemistry*.

Much of the work on CDO has been on investigating the Cys-Tyr crosslink, but no structures existed of non-crosslinked mammalian CDO variants and thus a structural explanation for how the crosslink impacts catalysis was not known. In this study 17 CDO crystal structures in the presence of Cys and inhibitors, ranging from 1.25 to 1.65 Å resolution and mostly of site-directed mutants, are used to define the role of an active site Cys-Tyr crosslink in catalysis as well as to determine the mechanisms of CDO inhibition by homocysteine and azide. Main conclusions are that a chloride ion binds to the active site iron in both the C93A and Y157F variants and

Cys does not bind as it does to wild type CDO, but largely coordinates the iron only through its thiolate, and not through the α-amino group. The crosslink is necessary for persulfenate formation in the crystalline enzyme, as no Cys-persulfenate formed. These results defines a key role for Tyr157 in Cys binding, through both positioning Cys in the active site and modulating the pKa of its α-amino group. The structures in the presence of homocysteine and azide revealed why the CDO-homocysteine complex is unproductive for catalysis, and how azide binds to the wild-type CDO, associating with the iron and the hydroxyl of Tyr157 in the crosslinked enzyme.

Chapter 5: "Crystal Structures of *Bacillus subtilis* Cysteine Dioxygenase and a *Ralstonia eutropha* Cysteine Dioxygenase Homolog" Camden M. Driggers, Steven J. Hartman, and P. Andrew Karplus. In preparation as a Structure Report for *Protein Science*

To better understand how an enzyme functions, it can be helpful to study the properties of distant homologs that conserve function. Fortuitously, the structures of two bacterial CDO homologs are known that were originally solved by structural genomics groups, but have not been reported in the literature. These structures are of particular interest as they represent "Gln-type" and "Arg-type" classes of bacterial CDO homologs. The main conclusions from this work were that those "Arg-type" bacterial CDOs bind Cys in a manner similar to mammalian CDO, and that the "Gln-type" CDO homologs have a different active site and are actually more closely related to 3-mercaptopropionate dioxygenases.

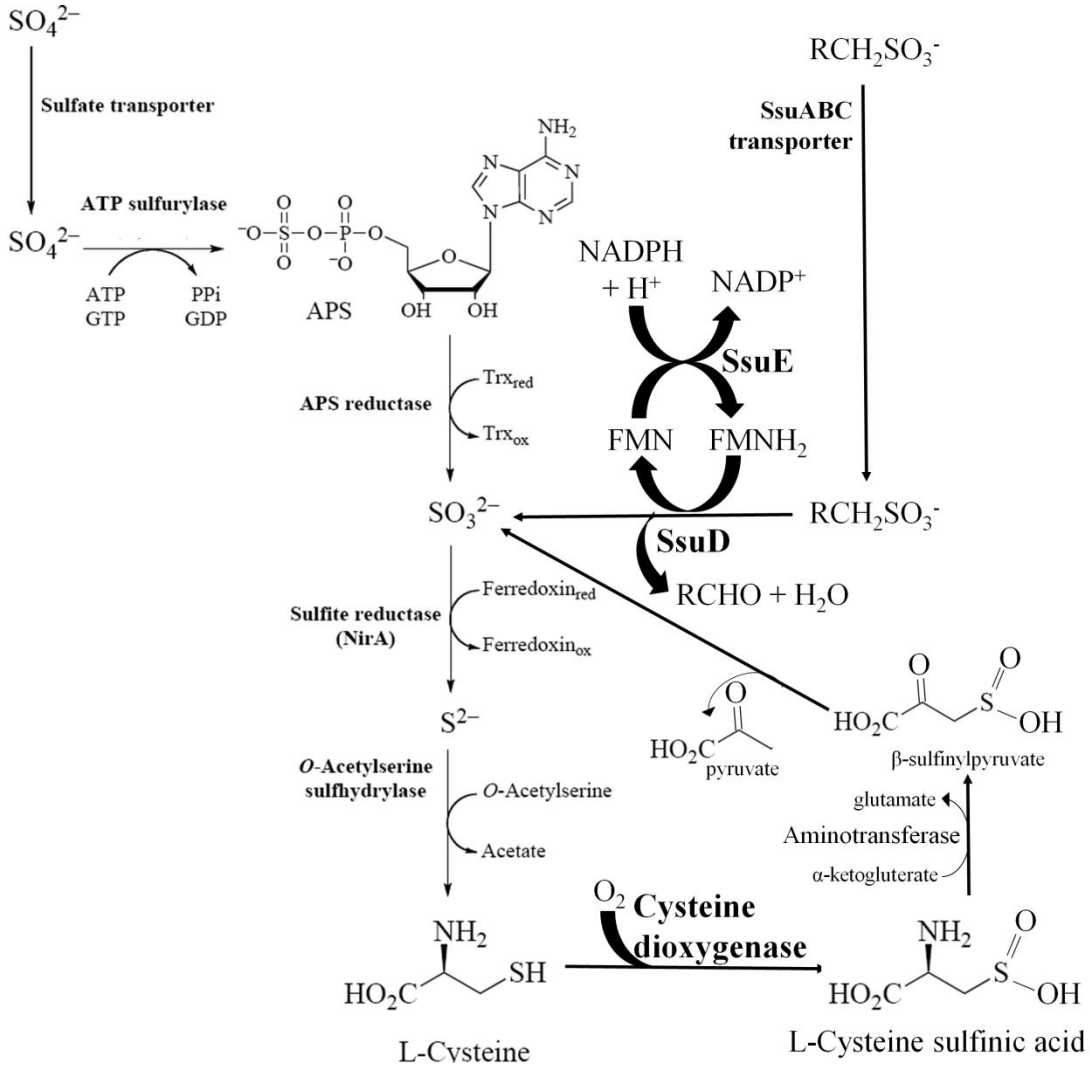


Figure 1.1. Two routes of sulfur assimilation to cysteine and one route of cysteine breakdown in *E.coli*. This Figure is modified from (6), and shows two alternative paths for sulfite acquisition, and a simple sulfur cycle that bacteria can use to maintain Cys levels by oxidizing Cys to CSA. When APS is in excess, the Ssu proteins are not expressed, when it is limiting they are expressed and *E. coli* is able to use alkanesulfonates to obtain sulfite.

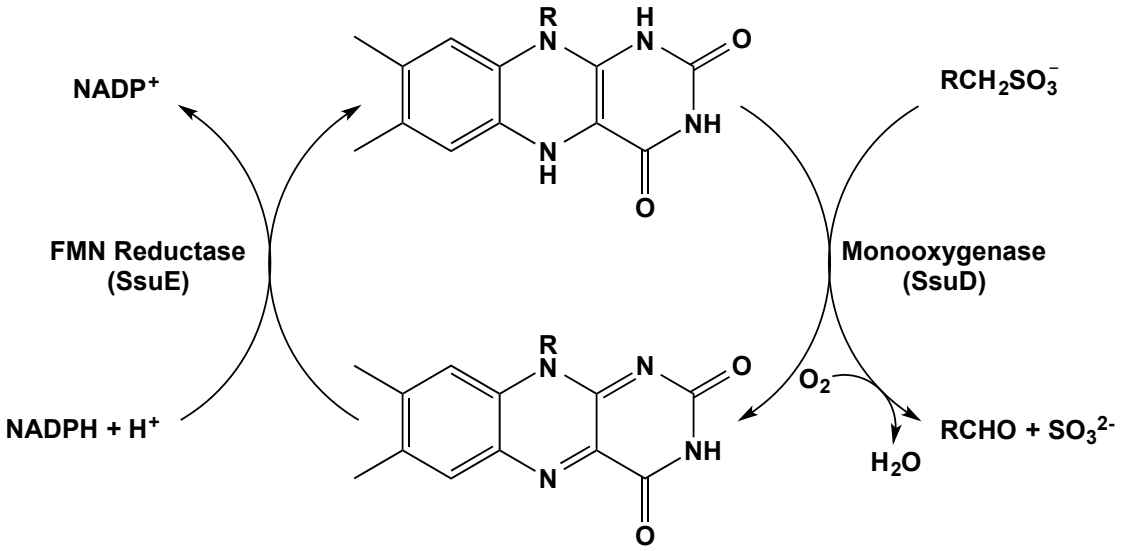


Figure 1.2. The reaction catalyzed by the SsuE/SsuD two-component monooxygenase system. This scheme illustrates that SsuE uses NADPH to reduce FMN, and then it supplies reduced $FMNH_2$ to SsuD. This kind of relationship between reductase and monooxygenase partners is seen for all two component monooxygenase systems. Flavin-reductases from at least three different enzyme superfamilies have evolved through convergent evolution, and the mechanisms of flavin reduction are different between the different superfamilies. Mechanism is typically conserved within a given superfamily, with SsuE being an apparent exception.

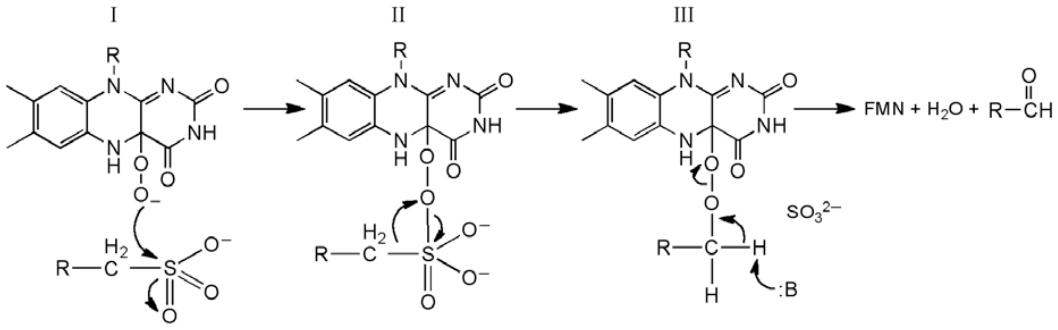


Figure 1.3. SsuD proposed mechanism (Figure 7 from (28)) The C4a-peroxyflavin makes a nucleophilic attack on the alkanesulfonate, cleaving sulfite for the bacteria to use.

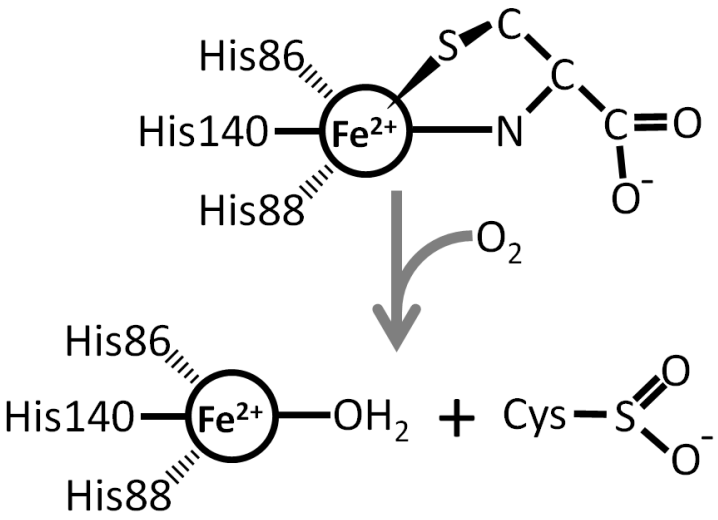


Figure 1.4. Reaction catalyzed by cysteine dioxygenase. This figure shows Cys binding first to the active site iron of cysteine dioxygenase, via its thiolate and α-amino group. After Cys binds then the oxygen binding site is made available and it binds, directly associating with the iron. From this point the mechanism is debated, and investigating the mechanism is the focus of our studies in Chapter 3.

References

1. Halle, B. (2004) Biomolecular cryocrystallography: structural changes during flash-cooling. Proc Natl Acad Sci U S A 101, 4793-4798

2. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. Science 336, 1030-1033

3. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38, W545-549

4. Shaker, M., Pascarelli, K. M., Plantinga, M. J., Love, M. A., Lazar, A. J., Ingram, D. R., von Mehren, M., Lev, D., Kipling, D., and Broccoli, D. (2014) Differential expression of cysteine dioxygenase 1 in complex karyotype liposarcomas. Biomark Cancer 6, 1-10

5. Brait, M., Ling, S., Nagpal, J. K., Chang, X., Park, H. L., Lee, J., Okamura, J., Yamashita, K., Sidransky, D., and Kim, M. S. (2012) Cysteine dioxygenase 1 is a tumor suppressor gene silenced by promoter methylation in multiple human cancers. PLoS One 7, e44951

6. Bhave, D. P., Muse, W. B., 3rd, and Carroll, K. S. (2007) Drug targets in mycobacterial sulfur metabolism. Infect Disord Drug Targets 7, 140-158

7. McCaddon, A., Hudson, P., Hill, D., Barber, J., Lloyd, A., Davies, G., and Regland, B. (2003) Alzheimer's disease and total plasma aminothiols. Biol Psychiat 53, 254-260

8. Kertesz, M. A. (2000) Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in Gram-negative bacteria. Fems Microbiol Rev 24, 135-175

9. Park, S., and Imlay, J. A. (2003) High levels of intracellular cysteine promote oxidative DNA damage by driving the Fenton reaction. Journal of Bacteriology 185, 1942-1950

10. Rocha, E. P., Sekowska, A., and Danchin, A. (2000) Sulphur islands in the Escherichia coli genome: markers of the cell's architecture? FEBS Lett 476, 8-11

11. Markovich, D. (2001) Physiological roles and regulation of mammalian sulfate transporters. Physiol Rev 81, 1499-1533

12. van der Ploeg, J. R., Eichhorn, E., and Leisinger, T. (2001) Sulfonate-sulfur metabolism and its regulation in Escherichia coli. Arch Microbiol 176, 1-8

13. Gao, B., and Ellis, H. R. (2007) Mechanism of flavin reduction in the alkanesulfonate monooxygenase system. Biochim Biophys Acta 1774, 359-367

14. Gao, B., and Ellis, H. R. (2005) Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. Biochem Biophys Res Commun 331, 1137-1145

15. Nissen, M. S., Youn, B., Knowles, B. D., Ballinger, J. W., Jun, S. Y., Belchik, S. M., Xun, L., and Kang, C. (2008) Crystal structures of NADH:FMN oxidoreductase (EmoB) at different stages of catalysis. J Biol Chem 283, 28710-28720

16. Schnell, R., Sandalova, T., Hellman, U., Lindqvist, Y., and Schneider, G. (2005) Siroheme- and [Fe4-S4]-dependent NirA from Mycobacterium tuberculosis is a sulfite reductase with a covalent Cys-Tyr bond in the active site. J Biol Chem 280, 27319-27328

17. Olney, J. W., Zorumski, C., Price, M. T., and Labruyere, J. (1990) L-Cysteine, a Bicarbonate-Sensitive Endogenous Excitotoxin. Science 248, 596-599

18. Stipanuk, M. H. (2004) Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annu Rev Nutr 24, 539-577

19. Ueki, I., Roman, H. B., Valli, A., Fieselmann, K., Lam, J., Peters, R., Hirschberger, L. L., and Stipanuk, M. H. (2011) Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize taurine and an increased catabolism of cysteine to hydrogen sulfide. Am J Physiol Endocrinol Metab 301, E668-684

20. Hogins, J., Crawford, D. C., Zorumski, C. F., and Mennerick, S. (2011) Excitotoxicity Triggered by Neurobasal Culture Medium. PLoS One 6

21. Heafield, M. T., Fearn, S., Steventon, G. B., Waring, R. H., Williams, A. C., and Sturman, S. G. (1990) Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. Neurosci Lett 110, 216-220

22. Stipanuk, M. H., Hirschberger, L. L., Londono, M. P., Cresenzi, C. L., and Yu, A. F. (2004) The ubiquitin-proteasome system is responsible for cysteine-responsive regulation of cysteine dioxygenase concentration in liver. Am J Physiol-Endoc M 286, E439-E448

23. Dominy, J. E., Jr., Hwang, J., Guo, S., Hirschberger, L. L., Zhang, S., and Stipanuk, M. H. (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is

regulated by substrate and represents a novel post-translational regulation of activity. J Biol Chem 283, 12188-12201

24. Dominy, J. E., Jr., Simmons, C. R., Karplus, P. A., Gehring, A. M., and Stipanuk, M. H. (2006) Identification and characterization of bacterial cysteine dioxygenases: a new route of cysteine degradation for eubacteria. J Bacteriol 188, 5561-5569

25. Stipanuk, M. H., Simmons, C. R., Karplus, P. A., and Dominy, J. E., Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. Amino Acids 41, 91-102

26. Chai, S. C., Bruyere, J. R., and Maroney, M. J. (2006) Probes of the catalytic site of cysteine dioxygenase. J Biol Chem 281, 15774-15779

27. Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Jr., Begley, T. P., Stipanuk, M. H., and Karplus, P. A. (2008) A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. Biochemistry 47, 11390-11392

28. Ellis, H. R. (2010) The FMN-dependent two-component monooxygenase systems. Arch Biochem Biophys 497, 1-12

Chapter 2

Crystal structure of *Escherichia coli* SsuE: Defining a General Catalytic Cycle for FMN reductases of the Flavodoxin-like Superfamily

Camden M. Driggers, Paritosh V. Dayal, Holly R. Ellis, and P. Andrew Karplus

Published in *Biochemistry, 2014, 53 (21), pp 3509–3519*

Abstract

The Escherichia coli sulfur starvation utilization (ssu) operon includes a two-component monooxygenase system made up of a nicotinamide adenine dinucleotide phosphate (NADPH)-dependent flavin mononucleotide (FMN) reductase, SsuE, and a monooxygenase, SsuD. SsuE is part of the flavodoxin-like superfamily, and we report here the crystal structures of its apo, FMN-bound and $FMNH_2$-bound forms at ~2 Å resolution. In the crystals, SsuE forms a tetramer that is a dimer of dimers similar to those of seen for homologous FMN-reductases, quinone reductases, and the WrbA family of enzymes. A π-helix present at the tetramer building interface is unique to the reductases from two component monooxygenase systems. Analytical ultracentrifugation studies of SsuE confirm a dimer-tetramer equilibrium exists in solution with FMN binding favoring the dimer. As the active site includes residues from both subunits, at least a dimeric association is required for the function of SsuE. The structures show that one FMN binds tightly in a deeply held site which makes available a second binding site, in which either a second FMN or the nicotinamide of NADPH can bind. The $FMNH_2$-bound structure shows subtle changes consistent with its weaker binding compared to FMN. Combining this information with published kinetics studies, we propose a general catalytic cycle for two-component reductases of the flavodoxin-like superfamily, by which the enzyme can potentially provide $FMNH_2$ to its partner monooxygenase by different routes depending on the FMN concentration and the presence of a partner monooxygenase.

Introduction

When starved for inorganic sulfur and cysteine, Escherichia coli can acquire sulfur from alkanesulfonates through the expression of the proteins of the ssuEADCB operon (1). This operon is transcriptionally induced (reviewed in 2), and encodes an ABC-type transporter specific for alkanesulfonates, an NADPH-dependent FMN-reductase (SsuE), and an $FMNH_2$-dependent alkanesulfonate monooxygenase (SsuD). The latter two enzymes make up a two-component monooxygenase system (Figure 2.1A) in which SsuE provides $FMNH_2$ to SsuD so it can carry out an oxygen dependent cleavage of the sulfonate group from compounds, such as C-2 to C-10 unsubstituted linear alkanesulfonates, substituted ethanesulfonic acids, and even sulfonated buffers (3-5).

The sulfur-starvation systems of *Pseudomonas putida* (6) and *P. aeruginosa* (7) are essentially the same as the *E. coli* SsuE/SsuD system, and species from at least 12 additional genus of bacteria (*Shigella, Citrobacter, Enterobacter, Klebsiella, Yokenella, Cronobacter, Erwinia, Pantoea, Dickeya, Brenneria, Serratia, and Yersinia*) have a close homolog (>60% sequence identity) to SsuE. For all such species with known genomes, the SsuE gene is clustered with that of an SsuD homolog. In some bacteria, such as *Bacillus subtilis* (8), *Burkholderia cenocepacia* (9), and *Corynebacterium glutamicum* (10), an SsuD homolog is in an operon lacking an SsuE-like reductase. In these organisms, SsuD may not require a specific $FMNH_2$ donor, and cases have been reported of flavin reductases associated with a particular two-component system being successfully used in an unrelated system (11, 12).

Two component flavin-dependent monooxygenase systems perform a wide variety of reactions (13), and although the various FMN reductases involved are sometimes mistakenly described as being evolutionarily related (e.g. see Figure 9 of 14), those identified to date actually are of three distinct fold types which, according to the SCOP database, belong to the ferredoxin reductase, the nitroreductase, and the flavodoxin-like superfamilies.

The reductases belonging to the ferredoxin-reductase and the nitroreductase super-families each have a well-defined catalytic cycle characteristic of their superfamily (13, 15). However, for two-component FMN reductases in the flavodoxin-like superfamily, SsuE and EmoB, the reductase component of a two-component system in Mesorhizobium sp. BNC1, are the only characterized examples and they are reported to have distinct mechanisms (13, 15). For SsuE, as summarized in a recent review (5), NADP+ inhibition studies (16) together with stopped-flow kinetics (17) and the 1000-fold tighter binding of FMN compared with $FMNH_2$, were taken to imply an ordered sequential reaction mechanism in which FMN is not a prosthetic group, but NADPH binds first, followed by FMN which is reduced and then released (Figure 2.1B). In contrast, for EmoB (14), kinetics studies and crystal structures have revealed an FMN-bound enzyme that transfers a hydride from NADH to a second FMN in a ping-pong type reaction mechanism (Figure 2.1C).

To further investigate the mechanism and properties of E. coli SsuE, we have determined crystal structures of the recombinant enzyme with no ligand bound, with FMN bound, and with $FMNH_2$ bound. These structures, together with comparisons with EmoB and other NAD(P)H-dependent FMN reductases in the flavodoxin-like superfamily, lead us to propose a revised mechanism for SsuE that is similar to that of EmoB and the other FMN reductases in this superfamily. We also present a general catalytic cycle that provides a rationale for the differences in the observed kinetics of SsuE and EmoB.

Methods

Protein purification and crystallization. Recombinant SsuE was purified as previously described (16) and stored in 25 mM potassium phosphate, pH 7.5, 10% glycerol, and 100 mM NaCl. For crystallization, SsuE was concentrated to ~10 mg/mL in 10 mM HEPES, pH 7.0. Building on previous work (18), crystals were grown at room temperature using hanging drops made with 4 μl of protein stock mixed with 2 μl of a reservoir containing 7.5% (w/v) polyethylene glycol (PEG) 3350 and 0.1 M

sodium citrate. For storage and soaks, crystals were transferred to an artificial mother liquor (AML) like the reservoir but with 20% (w/v) PEG 3350. For freezing crystals for data collection, an AML with 20% (w/v) glycerol (and any relevant ligands) was used as a cryoprotectant, and because the crystals were not fully stable in this solution, they were placed in it only briefly (<30 s) before flash freezing in liquid nitrogen.

Crystal soak and data collection. For preparing an FMN-bound complex, crystals were soaked in an AML containing 1 mM FMN. Crystals initially showed stress lines but re-annealed over ~12 h before being flash frozen. To obtain an $FMNH_2$ complex, equilibrated FMN-soaked crystals were added to a degassed solution containing 1 mM FMN, 30 mM dithionite with 50 μM methyl viologen. The blue color of reduced methyl viologen showed that the solution remained reduced during the <10 min soak before the crystals were flash frozen. NADP+ soaks used concentrations from 1 to 100 mM NADP+ either with or without 5 μM to 1 mM FMN and with or without 30 mM dithionite with 50 μM methyl viologen; additional attempts included presoaking crystals in 1 mM FMN and adding varying amounts of either NADPH or NADP+.

Data were collected at beamlines 5.0.1 and 5.0.3 at the Advanced Light Source (Lawrence Berkeley National Laboratory) as well as using our laboratory Rigaku RU-H3R rotating anode (Cu-Kα) and an R-Axis IV++ image plate detector. Synchrotron data sets were processed and scaled using the HKL2000 suite of programs (19). The RAXIS IV++ data set was processed using imosflm (20) and scaled using Scala (21) from the CCP4 program suite.

Structure determination and refinement. Full refinements were initially done for each dataset using resolution cutoffs based on $\langle I/\sigma \rangle \geq 2$ and $R_{meas} \leq 70\%$. Then, following recommendations showing the value of including weaker data (22), we reprocessed the data using a $CC_{1/2} \sim 0.2$ cutoff criterion, and extended the resolutions from 2.4 to 2.1 Å for the synchrotron apo-SsuE data, from 2.3 to 1.9 Å for the FMN-bound data, and from 2.6 to 2.3 Å for the $FMNH_2$-bound data (Table 2.1). For all refinements, the same random 5% of data were flagged for use in R_{free} calculations.

Using an initial 2.8 Å resolution laboratory based data set (data not shown), the crystal structure of apo-SsuE was solved by molecular replacement using Phaser (23); it placed four copies of a search model comprising chain A of the EmoB structure (PDB code 4LTD) with all waters removed. Model extension using the AutoBuild option in Phenix (24), led to an R/R_{free} of 33%/40%. Further refinements used Coot (25) for manual modeling, Refmac (26) and Buster (27) for minimizations, and Molprobity (28) to monitor the stereochemical quality of the models. TLS refinement was done defining each chain as one TLS group. When synchrotron data extending to 2.4 Å resolution became available (Table 2.1), they were used. During iterative manual model building, 268 water molecules were added in places having $2F_o$-F_c electron density $\geq 1\ \rho_{rms}$, and F_o-F_c density $\geq 3\ \rho_{rms}$, and that were within 3.7 Å from a potential hydrogen bonding partner. The final model refined using the 2.4 Å high-resolution cutoff had R/R_{free} 19.0%/24.1%. After extending the data to 2.1 Å resolution based on the $CC_{1/2} \sim 0.2$ cutoff criteria (22), refinement was continued using Phenix. The improved electron density maps allowed further waters to be built, and led to a lower final R/R_{free} of 18.0%/22.9% even at the higher 2.1 Å resolution limit (Table 2.1).

For FMN-bound SsuE, refinement at 2.3 Å resolution began with a partially refined model of apo-SsuE that after rigid-body refinement yielded an R/Rfree of 27.3%/32.1%. Restrained refinement dropped the R/R_{free} to 21.3/27.7%, and adding 1 TLS group per chain produced an R/R_{free} of 19.3/24.7%. Further refinement, including B-factor, angle and distance restraint weight optimization (to 0.3, 1.5, and 3.0, respectively), dropped R/R_{free} to 15.6/20.7%. Extending the data to the $CC_{1/2}$-based 1.9 Å resolution limit, further refinement using Phenix improved the model and expanded it from 226 to 380 waters. The final R/R_{free} were 17.3%/20.8% at the extended 1.9 Å resolution limit (Table 2.1), indicating much less over-fitting compared to the 2.3 Å refinement. Refinement of the $FMHN_2$-bound SsuE also began with a partially refined apo-SsuE model and using a similar process led to a 2.6 Å resolution model with 144 waters and R/R_{free} 18.9%/23.2%. Upon resolution extension to the $CC_{1/2}$-based 2.3 Å

limit, the number of waters grew to 177, and the R/R_{free} dropped to 17.4%/22.1% at the 2.3 Å limit (Table 2.1). Four, six and four total sidechains were stubbed in the final FMN-bound (residue number and chain:0D, 11C, 11A, 30D), the $FMNH_2$-bound (0D, 1D, 25C, 44D, 57B, 97D) and the apo structures (0D, 45D, 174B, 174D), respectively. Also the default geometry restraints for the FMN forced the entire flavin ring to be in one planar group; before the final rounds of refinement these were edited to allow for "butterfly" bending by defining two planes in which the N5, N10 and C1' atoms are in both planar groups.

Spectroscopic studies of crystals. UV/VIS spectra of crystals prepared identically with the crystals giving the FMN-bound and the $FMNH_2$ bound structures were collected at the National Synchrotron Light Source beamline X26C. To test for dichroic absorption bands, spectra were taken for all crystal orientations around a 360˚ rotation axis with kappa=0 and using a kappa offset of 90˚ (i.e. a crystal tilt of 45˚).

Structure-based analyses. Crystal packing interactions were analyzed using PISA (29). Structural comparisons and structure based sequence alignments were carried out on July 03, 2013 using a Dali search (30) of the Protein Data Bank (31). For generation of the phylogenetic tree, all proteins were selected that had a Z-score ≥15 and that were ≤90% identical to any other selected protein. Gaps were removed from the SsuE sequence in the Dali alignment and the format was converted to fasta by Jalview (32). The phylogenic tree was generated using Phyml (33). The names of the families within the flavodoxin-like superfamily are from the SCOP database (34). For the sequence alignment shown in Figure 2.2, gaps in the SsuE sequence were not removed.

Analytical ultracentrifugation. The oligomeric state of SsuE in the absence and presence of flavin was evaluated by sedimentation velocity on an Optima XL-A analytical ultracentrifuge (Beckman Instrument, Palo Alto, CA). Samples of SsuE were exchanged into 25 mM phosphate buffer, 1 mM EDTA, 0.15 M NaCl at pH 7.0 with an Amicon Ultra-4 with a 10 K molecular weight cutoff (Millipore). Three different concentrations of SsuE (12.5, 25 and 37 μM) with and without the addition

of an equimolar amount of FMN were loaded into double-sectored cells and equilibrated to 20 ˚C. Sedimentation data at 280 nm were collected at a rotor speed of 40,000 rpm and a radial step size of 0.003 cm. The partial specific volume of 0.7465 cm^3/g was calculated based on the amino acid composition. The buffer density of 1.005235 g/mL was determined using a DA-310M precision density meter (Mettler Toledo, Hightstown, NF) at 20 ˚C. Plots and fits of the data were generated with DCDT+ software using 12-18 consecutive scans for the analysis (version 2.40 from J.S. Philo). Normalized sedimentation coefficient distributions g(s*) were calculated using DCDT+ that utilizes a time derivative approach for sedimentation velocity analysis that includes Lamm equation models of the boundaries (35, 36). The g(s*) function was fit to a Gaussian curve to determine the concentration, sedimentation coefficients, and diffusion coefficients of a particular species. The curves were best fit to a single species model.

Results

Overall Structure

Recombinant SsuE crystallized readily in space group $P6_122$ as previously reported (18) and we solved its structure by molecular replacement using the EmoB crystal structure (37% sequence identity). Structures with reasonable R-factors and at resolutions near 2 Å (Table 2.1) were determined in three states: apo-SsuE as grown, an FMN-bound form resulting from a 1 mM FMN soak, and an $FMNH_2$-bound form resulting from dithionite treatment of FMN-soaked crystals. In all of these structures, there are four chains in the asymmetric unit and the protein forms tetramers with 222 symmetry, with chains A and B and chains C and D constituting pairs that make two different kinds of half tetramers (Figure 2.2A). In both cases, the full tetramer is created by a crystallographic 2-fold. The packing of the unit cell is such that the chains A and C are better fixed in the lattice, having for the FMN-bound structure, average B-factors of 44 and 48 $Å^2$ compared with 64 and 77 $Å^2$ for chains B and D. The models for chains A through D, include the N-terminal Met (residue 0) through residue 172,

172, 173, and 173, respectively, with reasonably well defined electron density. The remaining 18 or 19 residues to the C-terminus (residue 191) had no clear electron density and were not modeled. A few sidechains without associated electron density were not modeled beyond Cβ (see methods). Unless otherwise specified, descriptions provided are based on chain A, but are representative of all the chains.

In describing the overall structure, we will use the FMN-bound form as it is both the highest resolution structure (1.9 Å), and the one that is most informative about function. As expected, in this complex all chains have strong 2Fo-Fc electron density for one bound FMN (Figure 2.2B) that is well-ordered and nestled deeply in the protein at the same site that is occupied by FMN in flavodoxins. Also, a second less ordered FMN with interpretable electron density only for the flavin and possibly the phosphate is bound in chains A, B, and C (further discussed below). SsuE has the typical flavodoxin fold with five central parallel β-strands in the order β2-β1-β3-β4-β5 flanked by α-helices, with two helices (α2A and α2B) occurring between β1 and β2 (Figure 2.2C, D). Another notable secondary structure feature of SsuE is a π-helix (37), 2 hydrogen bonds in length, present in helix α4 (Figure 2.2C, D).

Evolutionary Relationships

In a Dali structural similarity search (30), all top hits were members of the flavodoxin-like superfamily. The two closest relatives were EmoB (pdb entry 4ltm, 37% sequence identity; 1.6 Å rmsd for 168 Cα-atoms) (14), and an uncharacterized oxidoreductase from Corynebacterium dipththeriae "3k1y" (pdb entry 3k1y; 29% identity; 2.7 Å rmsd for 160 Cα-atoms). In their respective crystals, both of these enzymes form equivalent tetramers to that formed by SsuE (Figure 2.3A). The 3k1y protein is apparently a reductase in a two component system as its gene is in a gene cluster encoding an ABC transporter homolog and an SsuD homolog. Including representative more distant relatives found in the Dali search, we created a phylogenetic tree of the flavodoxin-like superfamily (Figure 2.3B). The tree shows four major groupings which are the NAD(P)H:FMN reductases, the quinone reductases, the WrbA-like proteins and the flavodoxin-related proteins. In this tree,

SsuE, EmoB and the 3k1y protein form a sub-branch of the NADPH:FMN reductase group. We note that the π-helix associated with helix α4 and centered on Tyr 118 is only found in the proteins in the SsuE sub branch and is conserved among them. As is common for π-helices (37), this is consistent with it being evolutionarily derived by a single residue insertion that was associated with a gain of function related to the differentiation of the two-component reductase subgroup (see Figure 2.2D and highlight in Figure 2.3B).

Although no functionally characterized members of the NADPH:FMN reductase family outside of the SsuE sub-branch are part of a two-component system, all of them catalyze similar chemistry in that they have an FMN prosthetic group that, via a ping-pong type mechanism like that proposed for EmoB (Figure 2.1C), can be reduced by NAD(P)H and then reduce a variety of substrates such as quinones, azo dyes, chromate and cyanide (38-44). Also, all of them are dimers, tetramers, or proteins having a dimer-tetramer equilibrium and their geometries of association match the AB dimer and the tetramer of SsuE. As has been described (44), for tetramer formation there is some variation in the twist between the associating dimer pairs.

Quaternary Structure

The association in the crystal of SsuE into a tetramer that matches the physiologically relevant tetramers of related enzymes (Figure 2.3B) led us to suspect that it is a physiologically relevant association. This was surprising because we had previously observed that SsuE was a dimer in solution (1, 16). For this reason, we reinvestigated the oligomerization state of SsuE in solution using analytical ultracentrifugation. The data from 12-18 consecutive scans were used to obtain the g*(s) distribution at different SsuE concentrations. The data were best fit to a single species model as there was no other oligomeric species observed at the concentrations utilized in the analysis. The SsuE enzyme in the absence of flavin exhibited sedimentation coefficients of 4.6 S to give an average molecular mass of 73.1 ± 6.2 kDa (Figure 2.4). Given that the monomeric molecular weight of SsuE is 21.3 kDa,

the SsuE enzyme exists primarily as a tetramer in the absence of flavin. The SsuE enzyme used in the crystallization trials would predominately exist as a tetramer even before adding the crystallization reagents that might promote further association. However, in the presence of one equivalent of FMN, SsuE gave a sedimentation coefficient of 3.9 S and behaved as a pure dimer with an average molecular mass of 40.9 ± 1.7 kDa (Figure 2.4).

The dimer interface between chains A and B buries 1160 $Å^2$, and the additional surface buried when two dimer pairs come together to form a tetramer is ~2200 $Å^2$. In contrast, the largest crystal packing interfaces bury ~650 $Å^2$. One interesting feature at the tetramer interface is helix α4 and the π-helix common to SsuE, EmoB, and 3k1y (Figure 2.3B). This π-helix is located near the center of the tetramer forming interface (Figure 2.2A) where the hydroxyl of Tyr118, the conserved residue central to the π-helix, hydrogen bonds across the interface to Ala78-O in the β3-α3 loop. Because Ala78-N hydrogen bonds to O4 of the deeply-nestled FMN (Figure 2.5D) this provides a mechanism for how tetramer formation could be influenced by FMN binding. The loop just prior to this helix further links FMN binding with the tetramer interface, as the His112 sidechain and the backbone nitrogens of Thr109 and Gly108 hydrogen bond to the N3, O2 and N1 atoms of the tightly-bound flavin in the same subunit (Figure 2.5D).

Active Site

When apo SsuE crystals are soaked in FMN, the c-axis increases by ~2% (Table 2.1), suggesting some conformational changes occur. All chains have a deeply nestled FMN bound in strong density (Figure 2.2B, 2.5A) and chains A, B, and C also have a second FMN bound which is highly exposed with its flavin in rather weak electron density (Figure 2.5A) stacking on top of the flavin of the first FMN. For chain D, in which only the first FMN is present, the side chain of His149 from chain B of another tetramer in the crystal rearranges upon FMN binding and stacks above the first flavin blocking the second FMN binding site. This rearrangement of 147-152 loop in chain B appears to be the main cause of the unit cell change.

The interactions of the well-ordered FMN involve the whole ligand (Figure 2.5D). The phosphate is coordinated by Ser8, Ser13, Ser15, Arg10, an ordered water molecule which is hydrogen bonded to Tyr76 and Arg14-N. The ribityl is interacting with Thr106, Asp140, Arg14, Val75-O, and an ordered water molecule, and the flavin ring hydrogen bonds with the side chain of His112 (to N3), as well as the backbone atoms Lys77-N (to N5), Ala78-N (to O4), Gly108-N (to N1), and Thr109-N (to O2). Additionally, the binding of the deeply held flavin is influenced by interactions across the dimer interface, with the C7 methyl group just 3.2 Å away from Asp89'-O (the prime indicates a residue from the other subunit of the dimer), and with Lys85' hydrogen bonding to Tyr76-O and also to the Asp89' sidechain (Figure 2.5D). The flavin group is not perfectly planar, having butterfly bends (around the N1-N10 axis) ranging from 3˚ to 8.5˚ among the different chains.

The second FMN is rather loosely bound and the modeled positions of the ribityl and phosphate have very weak electron density reflecting some combination of relatively high disorder and low occupancy (Figure 2.5A). The only interactions of the putative phosphate and ribityl are with the Arg10 and Arg14 sidechains, and with the first FMN. The flavin moiety of the second FMN is interacting with the Lys77 amino group (H bond to O4) and is nicely stacked on the flavin from the first FMN with the N5 atoms just ~3.5 Å apart. Cross-dimer interactions are fairly extensive with Glu93' sitting above the flavin and residues 89'-91' stabilizing the Lys77 side chain position and a network of waters interacting with flavin atoms N3 and N5 (Figure 2.5D).

In apo SsuE, the FMN-binding site is occupied by about 10 ordered waters (Figure 2.5C). These waters roughly follow the positions to be filled by the FMN ribityl and phosphate groups, with two waters in place of the flavin moiety. The conformational changes associated with FMN binding include small movements of the backbone at residues 9-18, 74-79, 106-111, and 138-158. The biggest change is at residues 10-12, where the backbone shifts ~2 Å, and the sidechain of Arg10 moves ~ 8 Å to an 'in' position and becomes much better ordered.

FMN-soaked crystals were treated with dithionite to obtain an $FMNH_2$ complex. During this soak the crystals changed color to pale yellow indicating a successful reduction of the flavin. In addition, correlated UV/VIS spectra collected on oxidized and reduced SsuE crystals confirmed the presence of the targeted oxidation states (Figure 2.6). The major change upon FMN reduction is that only the deeply-seated FMN site is occupied in chains A, C, and D (Figure 2.5B), and no $FMNH_2$ is bound to chain B. Furthermore, the electron density for the bound $FMNH_2$ molecules is weaker, reflecting either lower occupancy or less rigidity. The lack of $FMNH_2$ binding in chain B may be explained in that Arg10 makes a crystal contact (with Ser141 of chain C') and this stabilizes the "out" conformation of Arg10 as seen in its lower temperature factors compared with the other chains. Apparently the weaker binding of FMNH2 (Kd=15.5 vs. 0.015 μM for FMN) (18, 45) is not strong enough to cause Arg10 to shift as happens upon FMN-binding. Interestingly, in chains A and C, Arg10 has partial occupancy of the 'in' and 'out' conformations modeled at 0.67 and 0.33, respectively. Based on the assumption that the 'in' conformation of Arg10 is tied to FMN-binding, we modeled these two FMNH2 ligands at occupancies of 0.67. In chain D, the crystal contact placing His148 above the flavin apparently provides additional stabilization of FMNH2 binding, as it has higher occupancy (modeled at 1.0 to match the "Arg10-in" conformation of that chain). The butterfly bending of the flavin ring upon treatment with dithionite, increases only slightly to range from 6˚ to 11˚ among the subunits.

Our attempts to obtain an NADP(H) complex (see methods) were not successful. NADP(H) soaks alone resulted in an apo structure and soaks with FMN and NADP(H) resulted in complexes looking like the FMN-complex. A NADP(H) complex was difficult to obtain because the second site is not made available until the first site is occupied, but FMN also competes with NADPH for the second binding site. The shared second binding site provides a structural explanation for the substrate inhibition effect on kinetics observed at high FMN concentrations (16). Interestingly, a published complex of EmoB with FMN and NADP (pdb entry 2vzj) which was

originally interpreted to place the nicotinamide next to the flavin (14), has been reinterpreted to show NADP binding in a non-productive mode with the adenine stacking on top of the first flavin (15; pdb entry 4ltn). An overlay of this structure with the dual FMN bound structures of EmoB and SsuE (Figure 2.7) shows that the plane of the adenine matches well the positions of the second FMN flavin. The NAD pyrophosphate group and nicotinamide extend in a different direction than the ribityl and phosphate of FMN. The only apparent hydrogen bond to this part of the NAD+ is between the proximal phosphate of NAD, and the C3' ribityl hydroxyl of the first FMN (Figure 2.7). The minimal interactions with the NMP half of the NAD are consistent with its high temperature factors (~60 $Å^2$ for the nicotinamide vs. ~30 $Å^2$ for the adenine) and weaker density. No crystal contacts appear to be influencing the NAD binding mode in that crystal.

Discussion

A prominent theme of these studies on E.coli SsuE is that this enzyme has tertiary and quaternary structural properties that are highly similar to EmoB and are also quite similar to the other reductase branches of the flavodoxin-like superfamily. All members of the flavodoxin-like superfamily bind FMN (or FAD for some quinone reductases) at a common site, which is the deeply nestled first site in SsuE. The flavodoxins themselves are largely monomeric, and for the one flavodoxin known to form a dimer in solution, the dimer is different than that of SsuE and the other flavodoxins (46). Flavodoxins only bind a single FMN, but bind it tightly whether it is in the reduced, semiquinone, or oxidized state. This relates to their function as pure electron transferases that cycle between the oxidized and semiquinone states (47), and is enabled by a conformational change in which the Gly57-Asp58 (numbering for Clostridium beijernickii (PDB 1fla; Figure 2.3B)) peptide bond flips upon flavin reduction, allowing Gly57-O to receive a hydrogen bond from the protonated N5 atom of the semiquinone or hydroquinone flavin, while Asp58-N donates a hydrogen bond to the lone pair of atom N5 of the oxidized flavin (e.g. 48). Also, the flavodoxin flavin

is protected from hydride transfer (i.e. 2 electron) reactions by having its upper (si) surface covered by a Tyr residue.

In contrast, for all of the other branches of the flavodoxin-like superfamily (the WrbA-like proteins, the quinone reductases, and the NAD(P)H-dependent FMN reductases; Figure 2.3B), neither the Gly that allows the peptide flip, nor the Tyr that covers the si-face of the flavin is present, making the second site above the flavin available for binding substrates for hydride transfer reactions. All of these enzymes that have been characterized have NAD(P)H:acceptor oxidoreductase activity, where the acceptor may be quinones or FMN or assorted other substrates (38, 40, 44, 49-52). Also, for all characterized members of these families except SsuE, the deeply held FMN binds as a prosthetic group and the catalytic cycle proceeds via the ping-pong type mechanism reported for EmoB (Figure 2.1C), with the nicotinamide reductant and the FMN, quinone, or other oxidant sequentially occupying the same second binding site.

Interestingly, the basic AB dimer structure seen in SsuE is conserved in all three superfamily branches that are reductases and in many cases the tetramer is also conserved. The dimer conservation appears related to function because the binding of the nicotinamide, flavin, or quinones in the second site is actually facilitated by the dimer mate. The WrbA-like members all form the same kind of tetramer in their crystals (Figure 2.3A), and either a tetramer or a dimer-tetramer equilibrium is reported to exist in solution (49, 53, 54). In one case FMN binding is reported to enhance tetramer formation (54), and in another case it has no effect (53). The quinone reductase members are dimers in the crystal, and those that have been characterized are also dimers in solution (50, 51).

The members of the NAD(P)H-dependent FMN reductase family are also dimers, tetramers or in a dimer-tetramer equilibrium, with one exception being an uncharacterized protein (PDB 3fvw; see results). For the enzymes which exist in a dimer-tetramer equilibrium, in some cases tetramer formation is reported to be facilitated by FMN binding (50, 51), but in others FMN binding has no effect on

quaternary structure (42, 43). SsuE was shown to exist as a tetramer in solution in the absence of flavin and a dimer in the presence of FMN, making it the only known case where FMN binding weakens tetramer formation. We do not have a good explanation for why earlier gel filtration studies done at pH=8 in the absence of flavin showed SsuE to be a dimer (3). We simply note that the conditions under which the quaternary structure is examined can, of course, influence behavior. For instance, a study at pH=4.0 using ESI-MS showed that the E. coli WrbA apoenzyme, has a dimer-tetramer equilibrium constant of >30 μM, and that concentrations of FMN higher than 25 μM promoted nearly complete tetramer formation (55). In contrast, studies on the same enzyme using analytical ultra-centrifugation at pH=7.9 showed a dimer-tetramer equilibrium constant of 1.4 μM with no change in the presence of FMN (43).

The fact that FMN weakens the tetramer of SsuE may relate to the unique features of SsuE, EmoB and 3k1y which all have the π-helix involving Tyr118 at the tetramer interface. Because the loop preceding it and the π-helix are close to the FMN binding site, it could link FMN-binding with tetramer dissociation. As these three enzymes represent a branch of the FMN-reductase family that interface with monooxygenases, it may be that the dissociation to dimers facilitates interaction with the partner monooxygenase.

Origins of lower affinity for $FMNH_2$

Another important question relates to the explanation for the 1000-fold higher affinity for oxidized FMN then for reduced $FMNH_2$ (16, 45). We find the most directly visible contribution for this is that SsuE does not undergo the peptide flip as seen for flavodoxins. Lys77 is the equivalent residue to the Gly in flavodoxin and is well held in place by hydrogen bonding by backbone and sidechains across the dimer interface (Figure 2.5D). In both oxidized and reduced $FMNH_2$, its backbone amide donates a hydrogen bond to the N5 of the oxidized flavin, and presumably upon flavin reduction forms an unfavorable electrostatic interaction with the protonated N5 atom. Although other specific interactions that disfavor $FMNH_2$ binding are not visible, the

lower occupancy of the $FMNH_2$ in the crystal compared with FMN is consistent with it having much weaker binding.

Lys77 is conserved only in the sub-branch that includes EmoB and 3k1y (Figure 2.3B), but in every single member of the WrbA-like, quinone reductase like, and NAD(P)H dependent FMN reductase-like families, the residue is not a Gly, and structural studies suggest that none of these enzymes undergo the peptide flip seen in flavodoxins.

Mechanistic implications

The high levels of similarity of SsuE with EmoB and with the broader reductases in the flavodoxin family make it surprising that SsuE would have a mechanism so distinct from them as is the proposed mechanism (seen in Figure 2.1B) with NADPH binding first and FMN binding second to make a ternary complex. Furthermore, given the structural features seen here, it is not plausible that NADPH binds before FMN, because the NADPH binding site seen in SsuE (and all of these enzymes) is not present until the first FMN is bound. The question then becomes, how can our earlier kinetics results supporting the mechanism shown in Figure 2.1B be reconciled with this?

In seeking an explanation, we compared the EmoB kinetics work (14), and our kinetics studies on SsuE (16), and discovered that an important difference in these studies is the concentration of FMN under which the studies were carried out. Although our previous inhibition studies were designed to be at "fixed non-saturating levels of FMN" (16), the studies were carried out at 0.01 μM SsuE and 0.10 μM FMN (16). These conditions would indeed be non-saturating for FMNH2 (<1% bound), but given the Kd for FMN, actually ~85% of SsuE would have FMN bound. Taking into consideration that the resting enzyme was not apo but ~85% saturated with FMN, the kinetics results showing that NADPH bound "first" would mean that the NADPH was actually binding to the FMN-bound enzyme rather than the apo enzyme. Thus the binding of FMN prior to NADPH is perfectly consistent with the steady-state kinetics data for SsuE (16). It is also perfectly consistent with our later rapid reaction kinetics

studies to determine the rate constants for individual steps in flavin reduction. These studies gave an initial fast phase associated with a charge-transfer complex between NADPH and FMN (17). The rate of this phase showed a hyperbolic dependence on the concentrations of both NADPH and FMN implying a rapid equilibrium binding of the two substrates. This observed rapid equilibrium binding of NADPH and FMN prior to the generation of charge transfer complex does not distinguish between pathways involving FMN binding first or NADPH binding first or a random order of binding. Thus the data are fully consistent with our structural results which imply that the formation of the FMN-bound SsuE complex enables the binding of NADPH.

This new interpretation, with NADPH associating with the FMN-bound form of the enzyme, is now similar to the behavior of EmoB. The main question remaining is whether the SsuE-bound $FMNH_2$ produced by hydride transfer from NADPH to FMN dissociates from the enzyme, or transfers the reducing equivalents to a second "substrate" FMN. In figure 2.8, we show a scheme for a general catalytic cycle that allows it to be seen that these two possibilities are simply special cases of a general mechanism that involves four possible steps: (1) FMN binds to the apo-enzyme; (2) NADPH binds to the FMN-bound enzyme, transfers a hydride to FMN, and dissociates as $NADP^+$; (3) a second FMN binds, is reduced by the first $FMNH_2$ and dissociates (or is transferred to the partner monooxygenase); (4) the first $FMNH_2$ dissociates (before a second FMN binds) and reforms the apoenzyme. The two possible steady state paths around the cycle are 1-2-4 if a single FMN is involved, or 2-3 if two FMNs are involved, one as a prosthetic group and the other as substrate. In the context of this model, the new interpretation is that SsuE follows the 1-2-4 route under conditions like those of our previous in-vitro studies (16) (Figure 2.8). However, in general which pathway is followed could vary depending on the affinities of a given enzyme for FMN and $FMNH_2$ as well as their concentrations.

Under the conditions of the EmoB kinetics studies (14), the enzyme ranges from 54% to 92% saturated with FMN, and we would suggest that some of the enzyme is following the 1-2-4 route while most of the enzyme follows the 2-3 route.

Such a mixture of paths could explain the relatively poor fit seen at low FMN concentrations (see Figure 6A of 14) when fitting the data to the strictly 2-3 mechanistic model. We suggest that in vitro SsuE and EmoB can proceed though either the 1-2-4 route or the 2-3 route depending on the solution conditions, and that what mixture of paths actually occurs in the bacteria is unknown. It might be even that the enzymes proceed through a different path at different times. Because the total concentration of FMN(H_2) for E. coli growing exponentially was ca. 50 μM (56), it seems that pathway 2-3 would be taken under these favorable growth conditions where FMN is abundant. However, under the starvation conditions that induce the expression of the SsuE/SsuD system, it is plausible that the enzyme typically proceeds through routes 1-2-4. Furthermore, this scheme also makes it clear how the preferred mechanism could easily be influenced by an association with a partner protein like SsuD.

Acknowledgements

We thank Dr. Dale Tronrud, Dr. Andrea Hall, and Dr. Russell Carpenter for useful scientific discussions and help with methods. We also thank Drs. Babak Andi and Allen Orville for collecting coordinated spectroscopic and diffraction data on reduced and oxidized FMN-bound crystals at beamline X26-C at the National Synchrotron Light Source (NSLS), supported by contracts DE-AC02-05CH11231 from the Office of Basic Energy Sciences of the U.S. Department of Energy. Synchrotron data were collected at the Advanced Light Source, supported by contract DE-AC02-98CH10886 from the Office of Basic Energy Sciences of the U.S. Department of Energy. This project was supported in part by Grant MCB-0545048 to HRE from the National Science Foundation.

Abbreviations

Sulfur starvation utilization components E and D, SsuE and SsuD; Ethylenediaminetetraacetic acid, EDTA; EDTA monooxygenase component B,

EmoB; tryptophan repressor binding protein, WrbA; 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, HEPES; polyethylene glycol, PEG; artificial mother liquor, AML; protein data bank, PDB; electrospray ionization mass spectrometry, ESI-MS; non-crystallographic symmetry, NCS.

Accession Numbers

Coordinates and structure factors for the SsuE models have been deposited in the Protein Data Bank with accession numbers as follows: SsuE in apo form (PDB code 4PTY); SsuE in oxidized FMN-bound form (PDB code 4PTZ); SsuE in reduced FMNH2-bound form (PDB code 4PU0).

Table 2.1. Data collection and Refinement Statistics for SsuE[a]

	Apo	**FMN-bound**	**$FMNH_2$-bound**
Data collection			
Unit Cell (Å)	a=b=185.8, c=91.6	a=b=186.7, c=93.4	a=b=187.0, c=91.3
Resolution (Å)	50-2.10 (2.14-2.10)	50-1.90 (1.93-1.90)	50-2.30 (2.34-2.30)
Unique Obs.	52483 (1922)	75282 (3711)	42033 (2051)
Multiplicity	9.5 (3.2)	30.6 (11.7)	19.4 (9.0)
Completeness	96.1 (71.1)	100 (100)	100 (98.5)
Average I/σ	18.2 (0.36)	46.1 (0.60)	30.7 (0.65)
R_{meas}[b] (%)	0.061 (>0.99)	0.087 (>0.99)	0.10 (>0.99)
R_{pim} (%)	0.021 (>0.99)	0.018 (0.998)	0.026 (0.644)
$CC_{1/2}$[c] (%)	1.00 (0.20)	1.0 (0.32)	1.00 (0.59)
Res <I/σ>~2 (Å)[d]	2.4	2.3	2.6
Refinement			
R_{cryst} / R_{free} (%)	18.0/22.9	*17.3/20.8*	17.4/22.1
No. residues	695	694	692
No. waters	275	380	177
No. atoms	5781	6124	5758
rmsd angles (°)	1.41	1.39	1.39
rmsd lengths (Å)	0.014	0.012	0.013
φ,ψ favored (%)[e]	96.7	98.1	95.6
φ,ψ outliers (%)[e]	0.15	0	0
<B> protein ($Å^2$)	61	58	71
<B> FMN 1/2($Å^2$)	-	43/77	67
<B> waters ($Å^2$)	55	56	67
PDB code	4PTY	4PTZ	4PU0

[a] All crystals had space group $P6_122$; Numbers in parentheses refer to highest resolution bin.
[b] R_{meas} is the multiplicity-weighted merging *R*-factor (*58*).
[c]$CC_{1/2}$ is the correlation between two datasets each based on half of the data as defined in Karplus & Diederichs (*22*).
[d]The resolution at which <I/σ> ~ 2 is given to allow comparison with previous standards for selecting high resolution limits.
[e]Ramachandran statistics as defined by Molprobity (*28*).

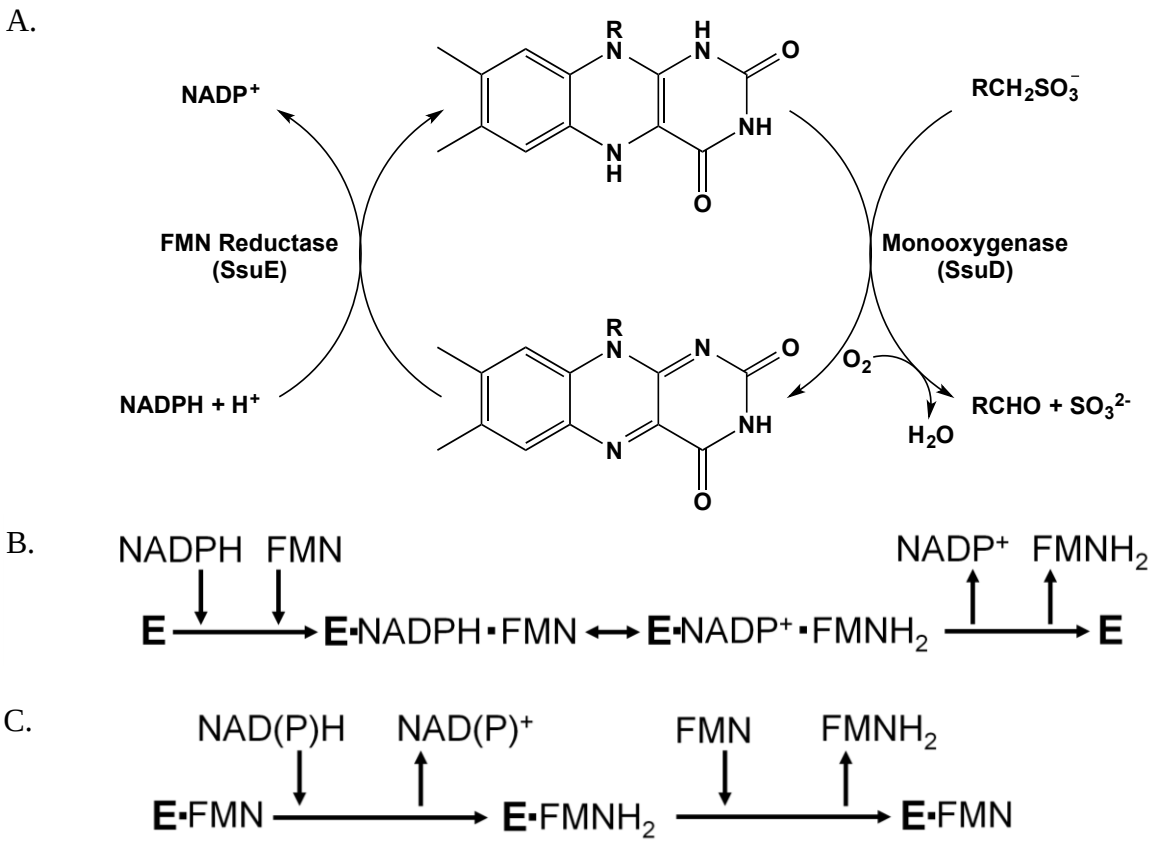
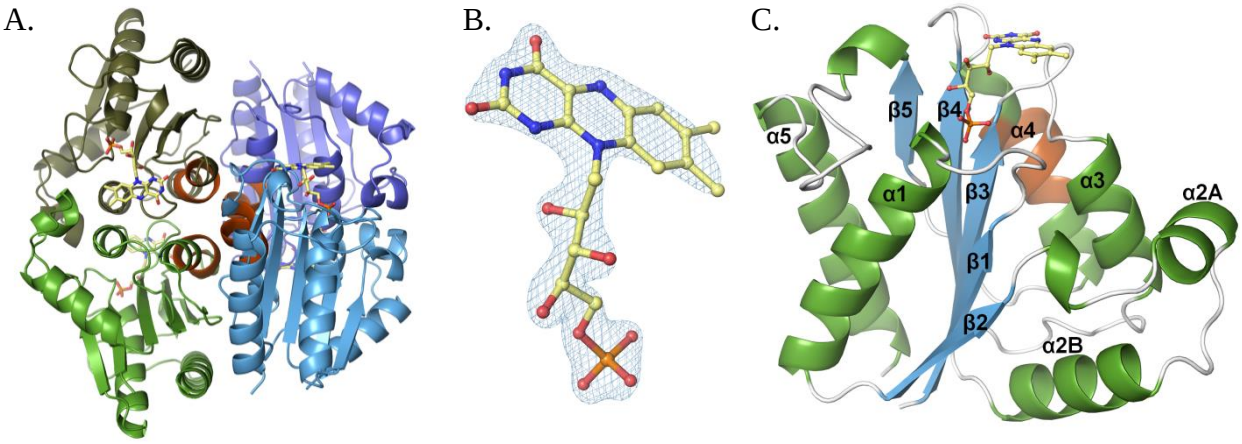


Figure 2.1. Reactions catalyzed by SsuE and other oxidoreductases of the flavodoxin-like superfamily. (A) The paired reactions catalyzed by the SsuE/SsuD two component system. (B) A sequential reaction mechanism proposed for SsuE (16). (C) Ping-pong bi-substrate bi-product reaction mechanism common to EmoB and other NAD(P)H-dependent FMN reductases of the flavodoxin-like superfamily.



D.

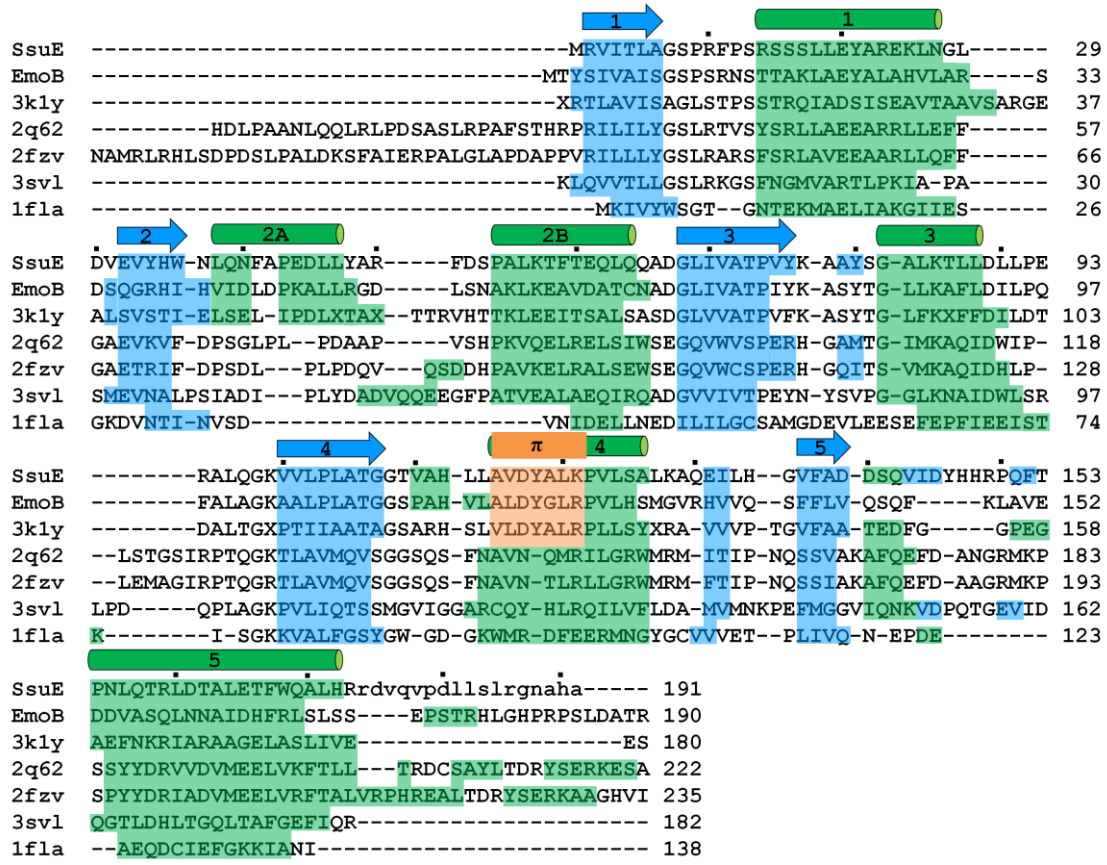
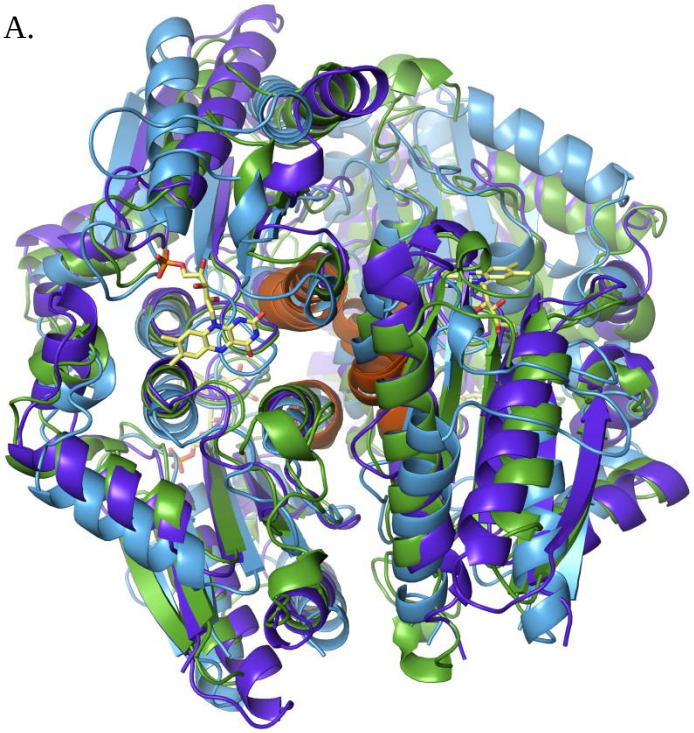


Figure 2.2. SsuE structure and structure-based sequence alignment with representative homologs. (A) SsuE tetramer made from the A/B chains (light/dark green) and the A'/B' chains (light/dark blue) created by the crystallographic two-fold. Also indicated

are the deeply nestled FMN (sticks with yellow carbons), and the π-helix (orange). The half tetramer formed by chains C/D is represented by the light green/light blue pair, and the basic dimer is represented by either the green A/B or the blue C/D pair. (B) $2F_o$-F_c electron density map (contoured at 2.0 ρ_{rms}) for the deeply-nestled FMN of chain A calculated after simulated annealing refinement was done with FMN removed from the model (i.e. an annealed omitmap). (C) Ribbon drawing of the SsuE monomer with colored and labeled secondary structures (β-strands –blue; α-helices –green; π-helix –orange). This view is rotated ~90˚ relative to the light-blue chain of panel A. (D) Structure-based sequence alignment with SsuE secondary structure segments annotated and colored as in panel C. Lower-case letters are for residues not built because of uninterpretable density. Shown are the sequences for SsuE, EmoB, an uncharacterized reductase from *Corynebacterium diphtheria* (3k1y), an NADPH-dependent FMN reductase from *Shigella flexneri* (2fzv; (*52*)), the chromate reductase from *E. coli* (3svl; (*39*)), and the G57D mutant flavodoxin from *Clostridium beijerinckii* but with wild-type sequence shown (1fla; (*48*)).

A.



B.

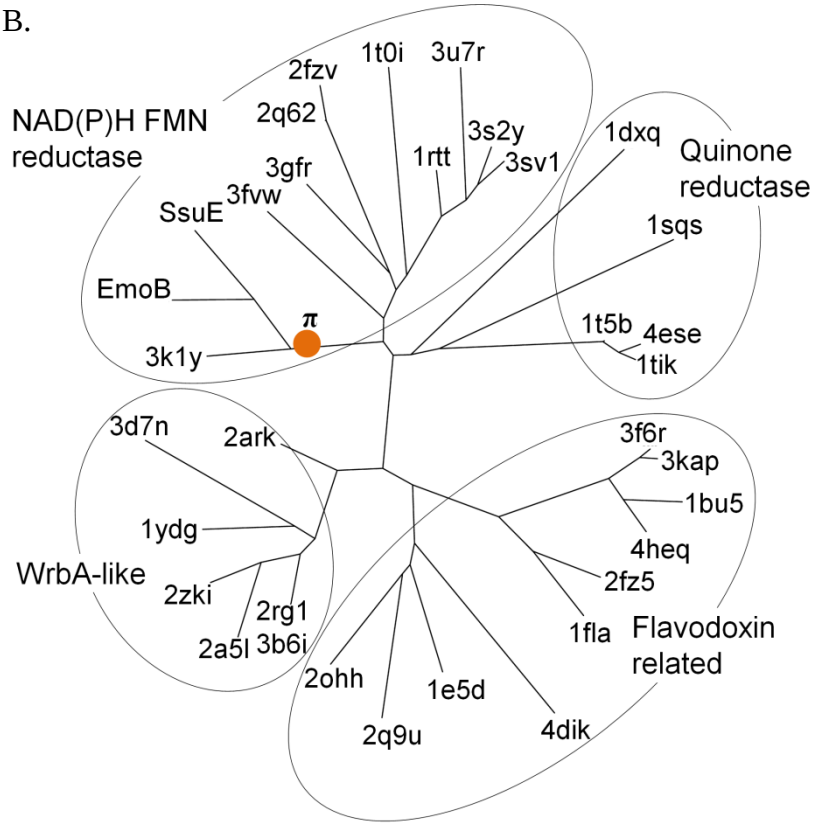

NAD(P)H FMN reductase
Quinone reductase
WrbA-like
Flavodoxin related
2fzv
1t0i
3u7r
2q62
3s2y
1dxq
1rtt
3sv1
3gfr
3fvw
SsuE
1sqs
EmoB
π
1t5b
4ese
3k1y
1tik
3d7n
2ark
3f6r
3kap
1bu5
1ydg
4heq
2fz5
2zki
2rg1
1fla
2a5l
3b6i
2ohh
1e5d
2q9u
4dik

Figure 2.3. Comparisons of SsuE with its homologs. (A) Overlay of the SsuE (blue), EmoB (green) and 3k1y (purple) tetramers with the deeply nestled FMN from SsuE shown (yellow carbons). The π-helix (orange) is located near the center of the tetramer in all three enzymes. (B) Phylogenic tree of the flavodoxin superfamily including proteins of known structure that aligned with SsuE with a Dali Z-score ≥15 and were ≤90% identical to all other hits (30). The phylogenic tree was generated using Phyml (33). The four families making up the flavodoxin-like superfamily are labeled using their SCOP classification (34). One gene product crystallized without FMN bound by a structural genomics group (pdb entry 3fvw), is unique among the NAD(P)H:FMN reductases in that it does not form a similar dimer or tetramer in the crystal.

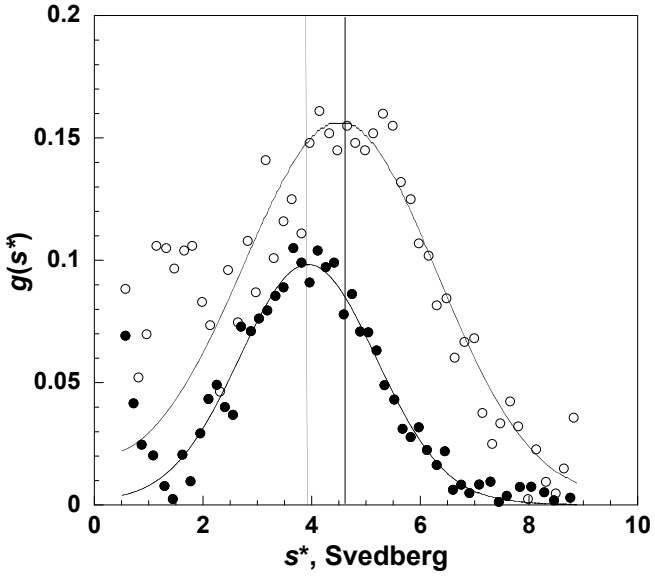


Figure 2.4. Sedimentation velocity studies of SsuE. Representative g(s*) distributions based on a single-species model are shown for SsuE (25 μM) in the absence (open circles) and presence of 25 μM FMN (closed circles).

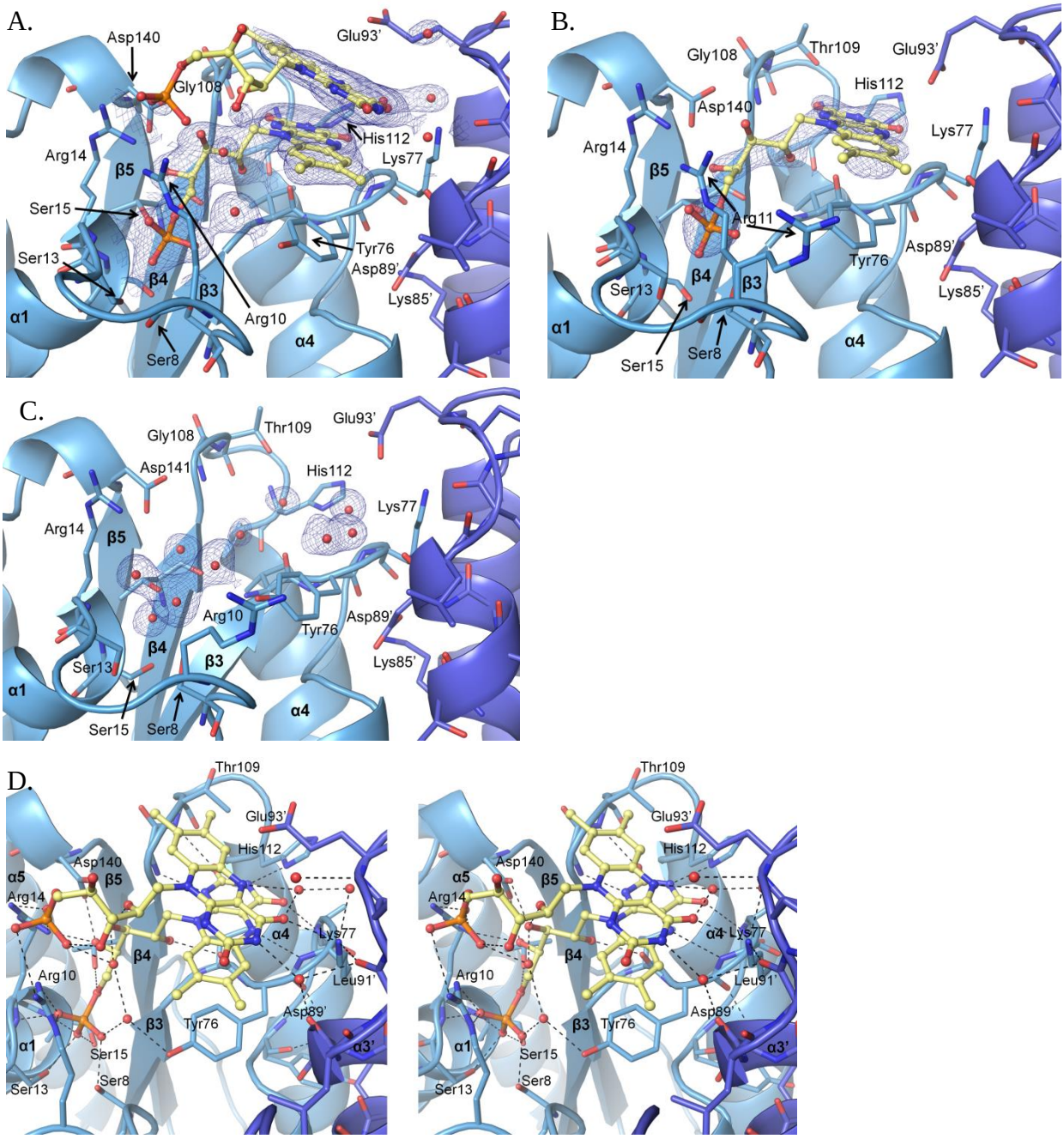


Figure 2.5. Three forms of the active site of SsuE. (A) FMN-bound SsuE is shown with chain A (light blue) and chain B (dark blue) and 2Fo-Fc annealed omitmap electron density for the bound FMNs contoured at 1.0 ρ_{rms}. (B) $FMNH_2$-bound SsuE shown as in panel A. (C) Apo SsuE with ordered waters in the active site shown as in panel A. (D) Stereoview of the environment and hydrogen bonding interactions of FMN-bound SsuE. Residues from the dimer mate (chain B) are marked with a prime symbol.

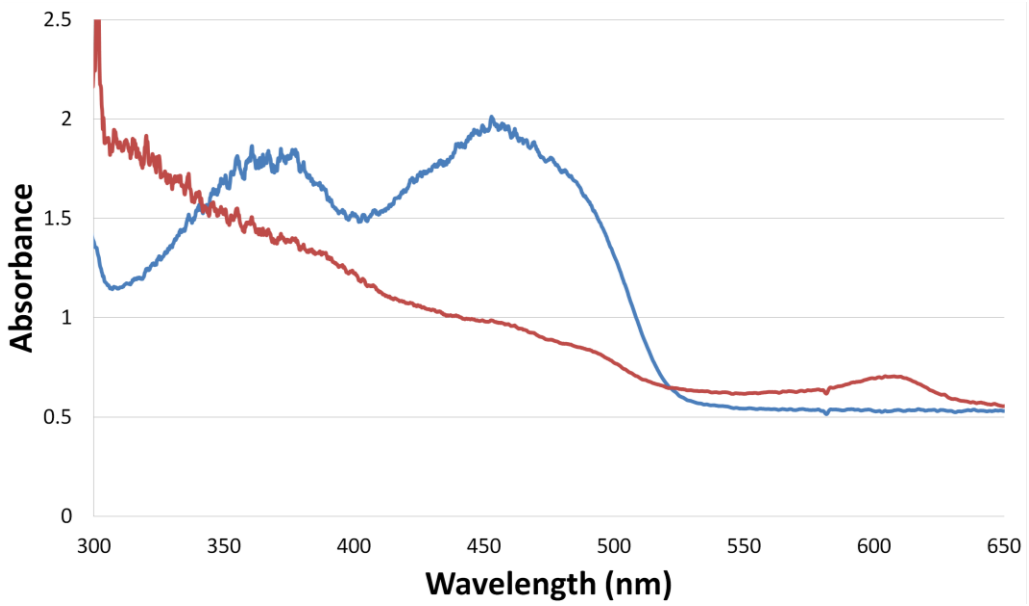


Figure 2.6. Spectra of SsuE crystals. UV/VIS spectra of an FMN-bound (blue) and an FMNH2-bound (red) SsuE crystal before data collection. Multiple crystals soaked in the same condition showed similar spectra. A constant (0.12) was subtracted from all of the values for the red spectrum, so that the two spectra had matching baselines. A small peak at 607 nm not characteristic of FMNH2 can be attributed to reduced methyl viologen present in the soak that produced the reduced crystal.

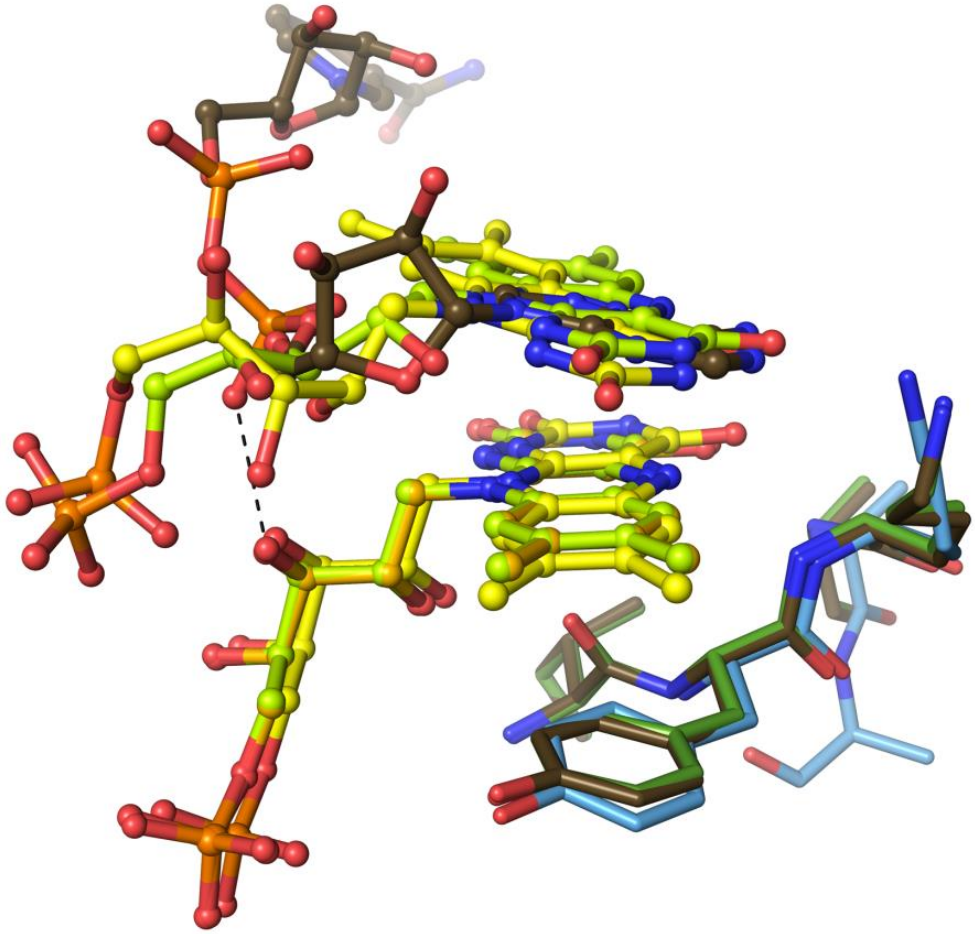


Figure 2.7. Comparing substrate binding in SsuE and EmoB. Overlay of the active site regions of dual FMN-bound SsuE (yellow carbons for FMN; light blue carbons for protein), dual FMN bound EmoB (light-green carbons for FMN; dark-green carbons for protein; PDB 4LTM), and the NAD and FMN bound EmoB models (orange carbons for FMN; brown carbons for NAD and protein; PDB 4LTN). Proteins were aligned using the CEalign feature implemented in Pymol (57).

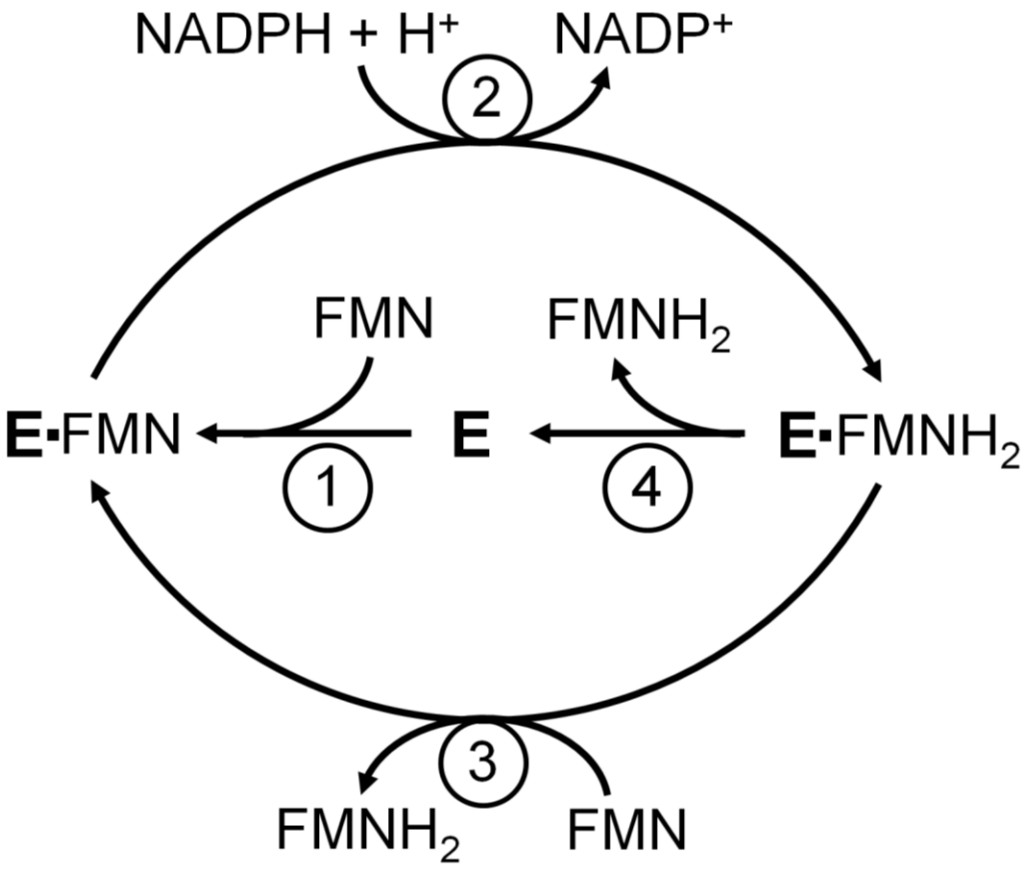


Figure 2.8. A general reaction cycle for the NADPH-dependent FMN reductases in the flavodoxin-like superfamily. When one FMN stays bound and acts as a prosthetic group, once primed by reaction 1, the enzyme cycles repeatedly though steps 2 and 3, and the second FMN acts as a substrate. When the first FMN is weakly bound in its FMNH2 form, and/or FMN is at relatively low concentrations, the first FMNH2 may dissociate before an FMN substrate can bind, leading the enzyme to cycle through steps 1, 2 and then 4.

References

1. van Der Ploeg, J. R., Iwanicka-Nowicka, R., Bykowski, T., Hryniewicz, M. M., and Leisinger, T. (1999) The Escherichia coli ssuEADCB gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. J Biol Chem 274, 29358-29365.

2. van der Ploeg, J. R., Eichhorn, E., and Leisinger, T. (2001) Sulfonate-sulfur metabolism and its regulation in Escherichia coli. Arch Microbiol 176, 1-8.

3. Eichhorn, E., van der Ploeg, J. R., and Leisinger, T. (1999) Characterization of a two-component alkanesulfonate monooxygenase from Escherichia coli. J Biol Chem 274, 26639-26646.

4. Xiong, J., and Ellis, H. R. (2012) Deletional studies to investigate the functional role of a dynamic loop region of alkanesulfonate monooxygenase. Biochim Biophys Acta 1824, 898-906.

5. Ellis, H. R. (2011) Mechanism for sulfur acquisition by the alkanesulfonate monooxygenase system. Bioorg Chem 39, 178-184.

6. Kahnert, A., Vermeij, P., Wietek, C., James, P., Leisinger, T., and Kertesz, M. A. (2000) The ssu locus plays a key role in organosulfur metabolism in Pseudomonas putida S-313. J Bacteriol 182, 2869-2878.

7. Quadroni, M., James, P., Dainese-Hatt, P., and Kertesz, M. A. (1999) Proteome mapping, mass spectrometric sequencing and reverse transcription-PCR for characterization of the sulfate starvation-induced response in Pseudomonas aeruginosa PAO1. Eur J Biochem 266, 986-996.

8. van der Ploeg, J. R., Barone, M., and Leisinger, T. (2001) Expression of the Bacillus subtilis sulphonate-sulphur utilization genes is regulated at the levels of transcription initiation and termination. Mol Microbiol 39, 1356-1365.

9. Iwanicka-Nowicka, R., Zielak, A., Cook, A. M., Thomas, M. S., and Hryniewicz, M. M. (2007) Regulation of sulfur assimilation pathways in Burkholderia cenocepacia: identification of transcription factors CysB and SsuR and their role in control of target genes. J Bacteriol 189, 1675-1688.

10. Koch, D. J., Ruckert, C., Rey, D. A., Mix, A., Puhler, A., and Kalinowski, J. (2005) Role of the ssu and seu genes of Corynebacterium glutamicum ATCC 13032 in

utilization of sulfonates and sulfonate esters as sulfur sources. Appl Environ Microbiol 71, 6104-6114.

11. Louie, T. M., Xie, X. S., and Xun, L. Y. (2003) Coordinated production and utilization of FADH(2) by NAD(P)H-flavin oxidoreductase and 4-hydroxyphenylacetate 3-monooxygenase. Biochemistry 42, 7509-7517.

12. Gisi, M. R., and Xun, L. Y. (2003) Characterization of chlorophenol 4-monooxygenase (TftD) and NADH : flavin adenine dinucleotide oxidoreductase (TftC) of Burkholderia cepacia AC1100. J Bacteriol 185, 2786-2792.

13. Ellis, H. R. (2010) The FMN-dependent two-component monooxygenase systems. Arch Biochem Biophys 497, 1-12.

14. Nissen, M. S., Youn, B., Knowles, B. D., Ballinger, J. W., Jun, S. Y., Belchik, S. M., Xun, L., and Kang, C. (2008) Crystal structures of NADH:FMN oxidoreductase (EmoB) at different stages of catalysis. J Biol Chem 283, 28710-28720.

15. Driggers, C. M., Ellis, H. R., and Karplus, P. A. (2012) Crystal Structure of Escherichia coli NADPH FMN reductase SsuE with and without bound FMN. Proceedings of the 17th International Symposium of Flavins and Flavoproteins. 17, 613-618.

16. Gao, B., and Ellis, H. R. (2005) Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. Biochem Biophys Res Commun 331, 1137-1145.

17. Gao, B., and Ellis, H. R. (2007) Mechanism of flavin reduction in the alkanesulfonate monooxygenase system. Biochim Biophys Acta 1774, 359-367.

18. Gao, B., Bertrand, A., Boles, W. H., Ellis, H. R., and Mallett, T. C. (2005) Crystallization and preliminary X-ray crystallographic studies of the alkanesulfonate FMN reductase from Escherichia coli. Acta Crystallogr Sect F 61, 837-840.

19. Minor, Z. O. a. W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol 276, p.307-326.

20. Leslie, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EAMCB newsletter on protein crystallography 26.

21. Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr D 62, 72-82.

22. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. Science 336, 1030-1033.

23. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J Appl Crystallogr 40, 658-674.

24. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D 66, 213-221.

25. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D 66, 486-501.

26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D 53, 240-255.

27. Bricogne, G. (1993) Direct Phase Determination by Entropy Maximisation and Likelihood Ranking: Status Report and Perspectives. Acta Crystallogr D 49, 37-60.

28. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D 66, 12-21.

29. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372, 774-797.

30. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38, W545-549.

31. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res 28, 235-242.

32. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189-1191.

33. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52, 696-704.

34. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) Scop - a Structural Classification of Proteins Database for the Investigation of Sequences and Structures. J Mol Biol 247, 536-540.

35. Stafford, W. F., 3rd. (1992) Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal Biochem 203, 295-301.

36. Philo, J. S. (2006) Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. Anal Biochem 354, 238-246.

37. Cooley, R. B., Arp, D. J., and Karplus, P. A. (2010) Evolutionary origin of a secondary structure: pi-helices as cryptic but widespread insertional variations of alpha-helices that enhance protein functionality. J Mol Biol 404, 232-246.

38. Gonzalez, C. F., Ackerley, D. F., Park, C. H., and Matin, A. (2003) A soluble flavoprotein contributes to chromate reduction and tolerance by Pseudomonas putida. Acta Biotechnol 23, 233-239.

39. Eswaramoorthy, S., Poulain, S., Hienerwadel, R., Bremond, N., Sylvester, M. D., Zhang, Y. B., Berthomieu, C., Van Der Lelie, D., and Matin, A. (2012) Crystal structure of ChrR--a quinone reductase with the capacity to reduce chromate. PLoS One 7, e36017.

40. Jin, H. J., Zhang, Y. F., Buchko, G. W., Varnum, S. M., Robinson, H., Squier, T. C., and Long, P. E. (2012) Structure Determination and Functional Analysis of a Chromate Reductase from Gluconacetobacter hansenii. PLoS One 7.

41. Agarwal, R., Bonanno, J. B., Burley, S. K., and Swaminathan, S. (2006) Structure determination of an FMN reductase from Pseudomonas aeruginosa PA01 using sulfur anomalous signal. Acta Crystallogr D 62, 383-391.

42. Liger, D., Graille, M., Zhou, C. Z., Leulliot, N., Quevillon-Cheruel, S., Blondeau, K., Janin, J., and van Tilbeurgh, H. (2004) Crystal structure and functional

characterization of yeast YLR011wp, an enzyme with NAD(P)H-FMN and ferric iron reductase activities. J Biol Chem 279, 34890-34897.

43. Grandori, R., Khalifah, P., Boice, J. A., Fairman, R., Giovanielli, K., and Carey, J. (1998) Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins. J Biol Chem 273, 20960-20966.

44. Binter, A., Staunig, N., Jelesarov, I., Lohner, K., Palfey, B. A., Deller, S., Gruber, K., and Macheroux, P. (2009) A single intersubunit salt bridge affects oligomerization and catalytic activity in a bacterial quinone reductase. FEBS J 276, 5263-5274.

45. Zhan, X., Carpenter, R. A., and Ellis, H. R. (2008) Catalytic importance of the substrate binding order for the FMNH2-dependent alkanesulfonate monooxygenase enzyme. Biochemistry 47, 2221-2230.

46. Hsieh, Y. C., Chia, T. S., Fun, H. K., and Chen, C. J. (2013) Crystal Structure of Dimeric Flavodoxin from Desulfovibrio gigas Suggests a Potential Binding Region for the Electron-Transferring Partner. Int J Mol Sci 14, 1667-1683.

47. Sancho, J. (2006) Flavodoxins: sequence, folding, binding, function and beyond. Cell Mol Life Sci 63, 855-864.

48. Ludwig, M. L., Pattridge, K. A., Metzger, A. L., Dixon, M. M., Eren, M., Feng, Y. C., and Swenson, R. P. (1997) Control of oxidation-reduction potentials in flavodoxin from Clostridium beijerinckii: The role of conformation changes. Biochemistry 36, 1259-1280.

49. Andrade, S. L. A., Patridge, E. V., Ferry, J. G., and Einsle, O. (2007) Crystal structure of the NADH : Quinone oxidoreductase WrbA from Escherichia coli. J Bacteriol 189, 9101-9107.

50. Faig, M., Bianchet, M. A., Talalay, P., Chen, S., Winski, S., Ross, D., and Amzel, L. M. (2000) Structures of recombinant human and mouse NAD(P)H : quinone oxidoreductases: Species comparison and structural changes with substrate binding and release. P Natl Acad Sci USA 97, 3177-3182.

51. Chen, H. Z., Hopper, S. L., and Cerniglia, C. E. (2005) Biochemical and molecular characterization of an azoreductase from Staphylococcus aureus, a tetrameric NADPH-dependent flavoprotein. Microbiol 151, 1433-1441.

52. Vorontsov, I. I., Minasov, G., Brunzelle, J. S., Shuvalova, L., Kiryukhina, O., Collart, F. R., and Anderson, W. F. (2007) Crystal structure of an apo form of Shigella flexneri ArsH protein with an NADPH-dependent FMN reductase activity. Prot Sci 16, 2483-2490.

53. Gorman, J., and Shapiro, L. (2005) Crystal structures of the tryptophan repressor binding protein WrbA and complexes with flavin mononucleotide. Prot Sci 14, 3004-3012.

54. Wolfova, J., Smatanova, I. K., Brynda, J., Mesters, J. R., Lapkouski, M., Kuty, M., Natalello, A., Chatterjee, N., Chern, S. Y., Ebbel, E., Ricci, A., Grandori, R., Ettrich, R., and Carey, J. (2009) Structural organization of WrbA in apo- and holoprotein crystals. Bba-Proteins Proteom 1794, 1288-1298.

55. Natalello, A., Doglia, S. M., Carey, J., and Grandori, R. (2007) Role of flavin mononucleotide in the thermostability and oligomerization of Escherichia coli stress-defense protein WrbA. Biochemistry 46, 543-553.

56. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol 5, 593-599.

57. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11, 739-747.

58. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4, 269-275.

Chapter 3

Cysteine dioxygenase structures from pH 4 to 9: Consistent Cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH

Camden M. Driggers, Richard B. Cooley, Banumathi Sankaran, Lawrence L. Hirschberger, Martha H. Stipanuk, and P. Andrew Karplus

Published in *Journal of Molecular Biology* (2013), 425(17), 3121-3136.

Abstract

Mammalian cysteine dioxygenase (CDO) is a mononuclear non-heme iron protein that catalyzes the conversion of cysteine (Cys) to cysteine sulfinic acid (CSA) by an unclarified mechanism. One structural study revealed a Cys-persulfenate (or Cys-persulfenic acid) formed in the active site, but quantum mechanical calculations have been used to support arguments that it is not an energetically feasible reaction intermediate. Here, we report a series of high-resolution structures of CDO soaked with Cys at pH values from 4 to 9. Cys binding is minimal at pH≤5 and persulfenate formation is consistently seen at pH values between 5.5 and 7. Also, a structure determined using laboratory-based X-ray diffraction shows that the persulfenate, with an apparent average O-O separation distance of ~1.8 Å is not an artifact of synchrotron radiation. At pH≥8, the active site iron shifts from 4- to 5-coordinate, and Cys soaks reveal a complex with Cys, but no dioxygen, bound. This 'Cys-only' complex differs in detail from a previously published 'Cys-only' complex which we reevaluate and conclude is not reliable. The high-resolution structures presented here do not resolve the CDO mechanism, but do imply that an iron-bound persulfenate (or persulfenic acid) is energetically accessible in the CDO active site, and that CDO active site chemistry in the crystals is influenced by protonation/deprotonation events with effective pKa values near ~5.5 and ~7.5 that influence Cys binding and oxygen binding/reactivity, respectively. Furthermore, this work provides reliable ligand-bound models for guiding future mechanistic considerations.

Introduction

Mammalian cysteine dioxygenase (CDO) is a non-heme iron protein that in its ferrous form (Fe(II)-CDO) catalyzes the conversion of cysteine (Cys) to cysteine sulfinic acid by incorporating both oxygen atoms of molecular oxygen to form the product (recently reviewed by Stipanuk et al(1)). The product from this first committed step in cysteine catabolism is then further catabolized to either taurine or pyruvate and sulfate. CDO expression is up-regulated in response to high cysteine levels,(2-4) which if not abated can be toxic.(5, 6) In general, tight regulation of CDO plays a key role in intracellular sulfur homeostasis.(3) Links between loss-of-function mutations in CDO and rheumatoid arthritis and some neurodegenerative disorders have been hypothesized based on observations of high cysteine and low sulfate levels.(5, 7, 8) In the CDO−/− mouse, taurine levels, but not sulfate levels, are low and metabolism of cysteine through alternative desulfhydration pathways is markedly elevated, giving rise to evidence of H2S toxicity.(6) CDO−/− mice exhibit a growth deficit, high postnatal mortality, joint hyperlaxity, connective tissue abnormalities including enlarged alveolar air spaces and abnormal organization of elastic fibers in the vasculature and parenchyma of lungs, and fatty acid oxidation defects.(6) Recently, CDO has been implicated as a novel tumor suppressor as decreased expression of CDO due to promoter methylation was seen in multiple tumor tissues, and the forced expression of CDO in cancer cells slowed their growth.(9)

Initial structural studies of recombinant rat CDO (92% identical to human CDO) showed it to have a cupin fold, with one structure revealing an unusual tetrahedral iron coordination involving three His residues and a water as ligands(10) and the other structure apparently containing nickel in place of the iron.(11) Both structures also revealed an unexpected thioether linkage between the Cys93 and Tyr157 side chains,(10, 11) which was later confirmed.(4, 12) This crosslink, not present in CDO from prokaryotes,(1, 13) is formed during enzyme turnover and is reported to occur in cells(4) and to increase catalytic efficiency 10-fold (4) or even more,(14) although its impact on activity and physiological relevance has been

recently questioned (15). In the unliganded enzyme, the Tyr157 hydroxyl forms a hydrogen bond with the iron-bound water and is part of a Ser153-His155-Tyr157 catalytic triad (10). A Tyr157Phe mutation leads to about a 50-fold loss in activity, consistent with the importance of this residue.(4, 16) Two structures of substrate-bound CDO have been reported: the first was a complex of human CDO with cysteine (16) and the second a complex of rat CDO with a putative cysteine persulfenate or persulfenic acid molecule (Figure 3.1A) formed in the active site during an aerobic soak with cysteine (17). In the persulfenate complex, Tyr157 was centrally located in close contact with both persulfenate oxygens, where it could serve as a catalytic acid/base (Figure 3.1A).

Taking the structural information together with the results of spectroscopic studies(18-20) and chemical considerations (21-23), there are now two main types of reaction mechanisms under consideration for CDO, both of which involve direct coordination of first Cys then molecular oxygen (18) to the active site iron. One mechanism (Figure 3.1B, left scheme), similar to that thought to occur for the uncatalyzed reaction and for Ni(II)-catalyzed thiol oxidations (17, 24-27), includes persulfenate formation which converts to product possibly via a thiadioxirane intermediate; in this mechanism the initial Cys oxidation occurs via sulfur attacking the Fe-proximal oxygen (OA). The other mechanism (Figure 3.1B, right scheme), more like that of monooxygenases, has the initial oxidation event involving transfer of the Fe-distal oxygen (OB) to generate Cys-sulfenate and an Fe(IV)-oxo intermediate. This latter mechanism has been supported by recent quantum mechanical calculations (22, 28), but not by a spectroscopic study implying that a Fe(III)-superoxo rather than a Fe(IV)-oxo intermediate facilitates substrate oxidation (19). The properties of small molecule models of the CDO active site are consistent with both mechanisms (29, 30). The challenge of defining the CDO mechanism has been exacerbated by the varied behaviors of recombinant CDO studied by different groups, as well as a lack of consensus in how to best assay CDO activity. The enzyme has been expressed in Escherichia coli using at least five different expression constructs, and the purified

proteins range from having 10% (11, 31) to 60% (19) iron bound that is from <1% (15) to 99% (18) in the ferrous form, and having amounts of the Cys93-Tyr157 crosslink ranging from minimal (15) to dominant (4). The activity assay we use, which includes 0.3 mM Fe^{2+} and a copper chelator and directly monitors the Cys-sulfinate product, gives consistent results for both natural rat liver CDO (2) and our recombinant enzyme (32), and an optimal pH range from ~5.5 to ~6.5. The only other published pH profile of recombinant enzyme did not use added iron, showed much lower activity, and an optimal pH range of ~7 to ~9 (33). Although other groups have shown good enzyme activity at pH 6.1 without added iron (18, 34), most current assays are carried out at pH 7.5 without added iron (11, 15, 18, 20, 33). A recent study showed that for Fe(III)-CDO, the inclusion of ascorbate as a reductant in assays can be used to recover enzymatic activity without adding ferrous iron (15). Also, even though Fe(III)-CDO is not enzymatically active (20, 35), spectroscopic titrations have shown Fe(III)-CDO can bind Cys between pH=5.5 and 9.5 (19). This implies that for CDO with iron in the ferrous form, this whole pH range (i.e. 5.5 to 9.5) could plausibly support enzyme activity.

In exploring how to gain additional information about the pH dependence of CDO properties, we found that the CDO crystal form we have studied (10, 17) is stable from pH=4 to pH=9 and provides a well-defined system – fully-loaded with iron and fully cross-linked – that can provide independent and direct information about the active site changes associated with variations in pH. Given this, plus the value of accurate structural information for informing mechanistic studies, here we follow up our initial brief report (10) of the putative persulfenate complex to describe it more fully, explore the pH dependence of its formation, and prove that it is not a synchrotron-induced artifact. At the higher pH values studied, a complex having only cysteine bound in the active site is observed. Because this complex has Cys bound differently than did the complex reported for human CDO (PDB code 2IC1) (16), we reevaluated the human CDO structure and found that structure does not actually represent a reliable Cys-bound complex.

Results and Discussion

Selecting the high resolution limits

This study encompasses nine structures in the main pH series of Cys soaks and four structures designed to provide specific additional insight. For these structures, the data quality and refinement statistics are given in Tables 3.1 and 3.2, respectively. Recently, Karplus & Diederichs (36) introduced a data quality indicator termed $CC_{1/2}$ and showed by way of a paired refinement strategy that improved models are obtained by extending the high resolution limit beyond those indicated by the current conventional standards of $I/\sigma \sim 2$ and $R_{meas} \sim 60\%$. Their main test case was one of the CDO data sets collected for this work (the Cys soak at pH=6.2), and they showed that improved models were obtained using data out to $CC_{1/2}$ between 0.1 and 0.2 even though the merging R-factors become very high. In order to create the best models for the structures in this study, we used such a $CC_{1/2}$-based cutoff throughout. As one additional test case, paired-refinements were done here for the pH=7.0 structure. As the resolution was extended, Rfree stayed the same, but the gap between Rwork and R_{free} decreased, indicating that the model was less over-fit when refinements included the weak high-resolution data (Table 3.3). Electron density maps also improved at the extended resolutions (e.g. Figure 3.2).

Conformational changes associated with Cys-persulfenate binding

Simmons et al. (2008) (17) was a brief report describing the observation of the Cys-persulfenate complex of CDO but did not include broader descriptions of conformational changes associated with ligand binding; we describe these more fully here. In the difference maps for persulfenate bound versus unliganded CDO (Figure 3.3A), the largest peak is due to the Fe atom shifting about 0.5 Å toward His86. The second strongest peak is for the sulfur atom of the bound Cys-persulfenate, and the whole Cys-persulfenate has strong positive features except for the OA (proximal) oxygen. The OA oxygen sits in negative density because its presence is more than offset by the loss of the tightly held iron-bound water that is displaced from nearly the same position upon substrate binding.

The Cys93/Tyr157 pair, Arg60 and Tyr58 all shift slightly (~0.3 Å) upon substrate binding and become more ordered in positions where they hydrogen bond with the persulfenate OA oxygen and α–amino group (Tyr157) or with its α-carboxylate (Arg60, Tyr58 and Tyr157) (Figure 3.3B). The Ser and His of the Ser-His-Tyr catalytic triad do not move as much as the Tyr, so the His-Tyr hydrogen bond elongates slightly from 2.7 Å to 2.9 Å. Additional difference map peaks distributed throughout the protein are associated with small global shifts in three lobes of the protein driven by the movement of the iron atom and the ligating His side chains. If the lobe including the iron ligands (residues 82-91, 100-124, 132-143, 161-179) is taken as a reference, the atoms at the edges of the other two lobes are shifted by ~1.0 Å (not shown).

pH rapidly equilibrates through the crystal

To probe how the formation or binding mode of the persulfenate ligand is influenced by the protonation state of active site residues, we solved the structures of CDO soaked with 100 mM Cys at pH values from 4.0 to 9.0. Evidence that the pH of the soaking buffer did equilibrate through the crystal was fortuitously provided by structural changes of residues 66-71 (Figure 3.4) and residue 79 which are both far from the active site. Residues 66-69 adopt a Type I` turn in all soaks at pH≤6.2, with the Asn67 side chain pointing outward to be near the His20 imidazole (Figure 3.4A). In contrast, for the pH=8.0 and pH=9.0 soaks, the backbone conformation shifts so that the Asn67-sidechain points inward and accepts an H-bond from the backbone amide of Lys69 (Figure 3.4B). The intermediate pH structures have a mix of the conformations, with ~60% of the high pH conformation seen at pH=7 (Figure 3.4C) and ~25% at pH=6.8. These differences are plausibly related to a deprotonation of His20.

For Glu79, at pH≤5.5 the side chain oxygens are close (~3.1 Å) to the backbone carbonyls of residues 147 and 148. However, in all soaks above pH=6.2, the side chain shifts to displace a water and form an H-bond with its own backbone amide. (In the pH=6.2 structure, only the high-pH conformation of Glu79 was modeled, even

though ~10% of the low pH conformation appears to still be present). The movement away from the electronegative carbonyl groups is consistent with the trigger being a deprotonation of the Glu79 to form the carboxylate. This Glu79 shift occurred even for a ~30 s soak at pH=4.5 (data not shown), indicating that pH equilibration through the crystal is quite rapid.

pH dependence of the persulfenate complex formation

pH range 5.5-7.0. In the active site, the persulfenate is formed at pH values ranging from 5.5-7.0 with minimal variations from the structure described by Simmons et al. (Figure 3.5D-G) (10). Compared to that structure, all of the structures presented here have a slightly higher occupancy of the bound persulfenate and the Cys93-Tyr157 crosslink, which we attribute to a minor difference between protein preparations. A notable feature is that in the pH=6.2, 6.8 and 7.0 structures (Figure 3.5E-G), the electron density for the persulfenate group is better resolved than was seen previously (17), and the observed persulfenate OA–OB distance is a rather long 1.7 – 1.9 Å (see further discussion below). One pH dependent change seen is that the Fe-S ligation distance complex varies among the persulfenate complexes: from 2.57 Å at pH=5.5, to 2.49 Å at pH=6.2, and ~2.38 Å at pH=6.8 and 7.0. These are all slightly longer than typical Fe-S bonds in high resolution protein structures (2.33 for S-Fe^{2+} and 2.29 for S-Fe^{3+}) (37). Since the resolutions of these structures are similar, we think the changes are not an artifact due to limited resolution of some data sets, and may represent some heterogeneity due to partial protonation of an active site group as pH decreases.

pH range 4.0-5.0. At pH=4.0, 4.5, and 5.0 (Figure 3.5A-C), the Cys soaks showed little to no ligand binding and, as was observed at pH=6.2 (10), a four-coordinate Fe with a solvent ligation distance of ~2.1 Å matching that expected for an Fe(II)-H_2O interaction (37). At these low pHs, Arg60 partially adopts a second conformation not compatible with Cys binding. At pH 4.0 and 4.5, difference map features indicate that about 20% of the molecules have lost iron with the iron-ligating His86 side chain moved away (Supplemental Figure A1.1). Because crystals soaked at

pH=4.0 for an hour no longer diffracted, that soak was done for only 20 min (Table 3.1). At pH=5.0, some weak density (1.5 ρ_{rms}) is present at the position where the Cys α-amino group coordinates the Fe (Figure 3.5C). We modeled this as a low-occupancy water, but it may reflect some binding of substrate. The lack of Cys binding in the crystals at pH≤5 is consistent with the lack of activity of CDO below pH=5 (2, 31, 32) as well as with the decreased binding of Cys to Fe(III)-CDO at acidic pHs (19). In the latter study, the decreased binding was attributed to protonation of the Cys α-amino and/or thiol group, and we agree that this is a reasonable assignment.

pH range 8.0-9.0. At the higher pH values of 8.0 and 9.0, Cys binds strongly, but no persulfenate is formed and there is also no binding of dioxygen (Figure 3.5H-I). Compared with the persulfenate complex at pH=6.2, the Fe-atom and Cys in these Cys-only complexes are shifted ~0.15 Å toward Arg60 with the Arg-carboxylate H-bonds shortening by about 0.2 Å (Figure 3.6). The bound Cys has additional slight shifts as the N- and S- coordination geometries adjust due to the lack of dioxygen. The Fe-S ligation distances at both pHs are 2.35 Å. Through a subtle domain motion, Tyr157 moves to accommodate the shift in the Cys (Figure 3.6). Interestingly, the new iron position is ~0.6 Å from that of the unliganded enzyme, meaning that the largest iron atom movement occurs upon Cys binding, with a smaller additional movement occurring upon dioxygen binding and persulfenate formation.

The natural conclusion would be that deprotonation of an active site residue, such as Tyr157, is responsible for the lack of oxygen binding and persulfenate formation at pH=8 and 9. However, because Fe(III)-CDO binds Cys but is not active (19, 20), we also considered whether the lack of persulfenate formation could be due to hydroxide facilitated oxidation of the active site iron – as can occur for free iron in aerobic solutions (38). To test this we collected data from crystals soaked with Cys at pH=8.0 in the presence of the reductant dithionite, and it also yielded a Cys-only complex (Figure 3.7B). A control Cys+dithionite soak at a lower pH yielded a bound Cys-persulfenate (data not shown), showing that even though dithionite is an oxygen scavenger, it does not sufficiently decrease oxygen levels to hinder persulfenate

formation during these long soaks. We also monitored UV-visible absorbance spectra on the pH=8 Cys-complex crystals, and it did not exhibit the 640 nm absorbance band associated with the Fe(III)-CDO-Cys complex (19, 20) and looked no different than that of the pH=6.2 Cys-persulfenate complex (Supplementary Figure A1.2).
We conclude that even if any iron oxidation occurred, it is minimal and it is not the cause of the lack of persulfenate. Given that Cys soaks yielded a different complex at pH=8 and 9 compared to pH=7, we also investigated if the coordination geometry of the unliganded active site changed at these pHs. The unliganded structure at pH=8.0 revealed a 5-coordinate iron with the fifth ligand modeled as a water (d ~ 2.3 Å) at a position equivalent to that occupied by the Cys α-amino group in the complex (Figure 3.7C). Although we cannot be confident what deprotonation event is causing the change in coordination, we note that Tyr157 is hydrogen-bonded to the water/hydroxide at the fourth coordination site and this ligand has a shorter coordination distance of ~1.95 Å (compared with 2.1 Å for the pH=6.2 unliganded enzyme), consistent with it being a hydroxide (Figure 3.7C) (37;39). These features are also seen in an equivalent soak at pH=9 (data not shown).

Laboratory X-ray source control data sets

It is known that protein structures solved at cryogenic temperatures using synchrotron radiation can display artifactual changes in structure due to both radiation damage (40, 41) and freezing (42, 43). To assess if the features seen in the unliganded or liganded CDO active site were significantly impacted by these factors, we used data from a rotating anode X-ray source to analyze a frozen low-temperature (LT) Cys-soaked crystal and also room temperature (RT) unliganded and Cys-soaked crystals. The LT control structure shows unequivocally that the persulfenate is not an artifact of synchrotron radiation. Indeed, the laboratory-based LT view of the Cys persulfenate complex is virtually identical to those based on synchrotron data (Figure 3.8A), including the observed OA–OB distance of ~1.8 Å.

We describe this ~1.8 Å OA-OB separation as a distance rather than a bond length, because the electron density with the apparent ~1.8 Å distance could represent

a mixture of structures having a range of OA-OB distances and geometries and levels of disorder. Structures that could be present are, for instance, Cys plus unreacted dioxygen, a Cys-persulfenate, a Cys-persulfenic acid, a Cys-sulfenate (Cys-SOH) plus a very weakly bonded oxygen, and even a thiadioxirane form. It is, however, of worth noting that for this unusual compound, the bond may actually be rather long, and indeed in the quantum mechanical calculations modeling a persulfenate intermediate (22), the optimized OA-OB bond was a nearly non-bonded 2.3 Å distance (and the Fe-S bond also lengthened to 2.42 Å). Kumar et al. interpreted these results to imply that a persulfenate species would not be stable in the active site and that "it seems likely the experimentally found 'persulfenate' structure actually refers to a protonated species" (22). As we already have noted, crystallography at this resolution cannot distinguish a protonated versus a deprotonated species, and it was an oversight of our original report (17) to not make it clear that the bound intermediate may be a persulfenate or a persulfenic acid. Indeed, a protonated neutral form could be favored by the OB environment, which is a weakly-polar pocket loosely bounded by Leu95-Cδ1, C93-Sγ, His140-Cε1, Y157-OH and His155-Cε1.

Regarding possible temperature-dependent changes, the unliganded RT structure at 2.15 Å resolution (Table 3.1) has several side chains with increased disorder as expected (42, 43), but the Fe coordination is essentially unchanged (Figure 3.8B). The most notable conformational change involves the active site residue Arg60 (and the nearby Met179) which mostly shifts to an alternate position (pointing toward the iron) that was weakly occupied in the low pH soaks (Figure 3.5A-C) and fully occupied in the structure with Cys164 modified by disulfide formation with Cys (see Figure 3.9C,D) (10;12). The RT Cys soak was not informative, because due to the long data collection time (days), it yielded a structure that had a perturbed active site because of disulfide formation at Cys164 (data not shown).

Re-examination of structure reported for human CDO in complex with Cys

The consistency of the binding mode of Cys across a wide pH range with and without persulfenate formation leads us to hypothesize that these structures are giving

us a reliable view of the physiologically relevant binding mode of Cys to CDO. Surprisingly, the binding mode seen here differs in key details from that reported in the 2.7 Å resolution analysis of Cys binding at pH=6.5 to human CDO (PDB entry 2IC1) (16). Three striking differences (Figure 3.9A) are the Arg60 side chain position, the cysteine Sγ-atom placement which in PDB entry 2IC1 matches the position of the iron bound water seen in unliganded structures, and the iron atom position which in the 2IC1 model also matches that seen in unliganded CDO structures. Interestingly, the 2IC1 model also has steric problems with bad contacts between the Cys carboxylate oxygens and both Arg60 (2.27 Å) and the Cys α-amino group (2.45 Å), as well as an implausibly short Fe-S ligation distance of just 2.02 Å (Figure 3.9B). Upon further inspection, we noticed that the positions of Arg60 and Met79 (which also differs from our Cys-only structure) match those that were previously shown in a high resolution analysis (10) to be associated with an artifactual disulfide formation at Cys164. Indeed, Ye et al. (16) comment on (and show in their Figure 3A) a substantial modification of Cys164 that they left uninterpreted.

Given these observations, we explored if the active site electron density from the 2IC1 data might be better explained by a modification of Cys164 rather than the binding of Cys as a substrate. An inspection of the electron density maps for entry 2IC1 shows that while the density for the modeled Arg60 and Met179 positions is strong, the density for the modeled Cys is quite weak for the Cα, Cβ, and N-atoms (Figure 3.9B). Also notable, is that the strong density into which the Cys Sγ-atom is modeled should not be taken as evidence for Cys-binding, because it sits at the same position that is occupied by the coordinating water of unliganded CDO. Furthermore, the maximal density of the iron peak indicates that the iron should be slightly shifted toward the Sγ-atom, which would make that already too short coordination distance even shorter. These observations support the conclusion that the electron density the Sγ-atom has been built into is actually due to the coordinating water present in unliganded CDO rather than a sulfur atom.

At the same time, the electron density is remarkably similar to that of the Cys164 modified enzyme of Simmons et al. (10) when those data are truncated to 2.7 Å (Figure 3.9C versus D). This similarity shows that all of the active site features of the 2IC1 structure can be explained by rearranged side chains and water molecules associated with Cys164 modification. Based on these observations, we conclude that 2IC1 is not a Cys complex.

Implications for understanding the mechanism of CDO

We undertook these studies to better understand the mechanism of CDO and related dioxygenases. The most significant results are the capturing of a Cys-only complex at pH 8 and above, and demonstrating that the CDO:persulfenate complex, until now seen only in a single crystal structure, is a robust result that is not an artifact of synchrotron radiation, and is consistently formed in the active site of CDO between pH 5.5 and 7.0, corresponding roughly with the pH range over which CDO shows activity in our standard assay done in the presence of 0.3 mM Fe^{2+} (32). As representatives of the complexes have high occupancies and are determined at high (1.25-1.4 Å) resolution, the conformations of the Cys-only and Cys-persulfenate ligands and their interactions with the active site residues have coordinate uncertainties of ~0.05-0.10 Å for the C, N, and O atoms and even better for the S and Fe atoms (see methods).

From the pH series, we conclude that Cys binding requires the deprotonated form of a group with an apparent pKa near 5.5, and that the coordination of the active site iron and the oxygen binding and reactivity of the Cys complex are both influenced by the protonation state of a group with an apparent pKa near 7.5. Unfortunately, crystal structures even at a resolution of ~1.4 Å do not show hydrogen positions, so without obvious changes in interactions (which we do not have here), we are not able to infer what groups are becoming protonated/deprotonated. As mentioned above, we postulate that the group with an apparent pKa near 5.5 is the Fe-bound Cys α-amino and/or thiol group, and that the group becoming deprotonated near pH=7.5 is the coordinating water in the unliganded complex and is Tyr157-OH in the Cys-bound

complex. As previously described (17), Tyr157 is well placed to interact with either of the persulfenate oxygens (OA and OB in Figure 3.1B) and serve as a catalytic acid/base during catalysis, and a pKa value near 7 or 8 is reasonable for a Tyr side chain acting as an acid/base catalyst (e.g. 44, 45).

Unfortunately, correlating our structurally-defined effective pKa values with those seen to influence the activity of CDO is difficult, because as noted in the introduction there is not agreement on the pH dependence of activity. Based on the pH-profile showing activity limited by apparent pKa values of 5.5 and 6.5 (2, 4, 32), we would infer that the deprotonation occurring in the crystal near pH=7.5 is blocking catalysis. However, based on the pH-profile with activity occurring between pH=7 and 9 (33), we would infer that the deprotonation event was generating the active form of the enzyme. Clearly, coming to agreement on how to best prepare and assay CDO and obtaining an accurate pH profile of its k_{cat} and k_M values are important next steps in helping define the mechanism of CDO.

Our discovery that the reported complex of human CDO with bound Cys (16) is not correct is also significant because that model served as the basis of quantum mechanical studies of the CDO mechanism (21-23, 28). Geometry optimization done in setting up the recent quantum mechanics study did correct the chemically unreasonable 2.0 Å S-Fe bond length to be ~2.3 Å (see figures 5 of Kumar et al.; 22), but other features of the 'optimized geometry' active site structure in the simulations of the persulfenate-based mechanism are different from the interactions consistently seen in the crystal structures presented here. Prominent examples are the interaction of Arg60 with the Cys α-carboxylate and the relative positions of the persulfenate oxygens and Tyr157 (see figure 11 of Kumar et al.; 22). Additional evidence for the critical nature of Tyr157 is that the residue is conserved in all CDO sequences (13) and is sitting in much the same position even in a bacterial CDO structure having <20% sequence identity with rat CDO (PDB entry 2GM6). Since catalytic proficiency can depend exquisitely on the precise placement of active site residues, it will be of

interest for some calculations to be guided more closely by the positions of the protein side chains seen here and in other reliable CDO structures.

As we see it, despite ambiguity about the protonation state of the persulfenate/persulfenic complex and uncertainty as to its relevance to the catalytic mechanism, the ability of the CDO active site to make and stabilize it proves that its formation is energetically accessible in the CDO active site. That an intrinsically less-stable molecule like the persulfenate rather than substrate or product would be stably trapped in the active site of the crystalline enzyme can be explained in that the on-enzyme equilibrium between substrate, various on and off-pathway intermediates, and product can be highly perturbed from that which would occur in solution. Also important to note is that these equilibria would also be temperature-dependent, so that the dominant form present in the active site may be influenced by the freezing process.

In considering the relevance to catalysis of this complex, one piece of evidence in favor of it being an intermediate would be to show that the crystalline enzyme is active. Unfortunately, our attempts thus far to show that crystalline CDO can convert Cys to the Cys-sulfinate product have not been successful (RB Cooley, unpublished). This could be due to a variety of factors, including something as simple as the high ionic strength and/or high PEG concentrations of the mother liquor causing technical challenges with the analyses (which they do), or something as subtle as the crystal lattice hindering certain dynamic modes of the enzyme that are required for complete turnover. Given the principle that active sites are optimized for stabilizing the transition state of a reaction, we think that this complex is closely related to an intermediate of the true catalytic mechanism even if it is not directly on-pathway. As was noted in a recent review (30), the proposal that CDO operates by monooxygenase like chemistry with the involvement of an Fe(IV)=O intermediate (Figure 3.1B, right path) is associated with a conceptual difficulty: why would such a powerful oxidizing agent as Fe(IV)=O be generated for such a relatively easy oxidation reaction? The mechanism of CDO has been surprisingly challenging to elucidate, but the accurate structures described here together with the insight into protonation/deprotonation

events influencing CDO chemistry provide a solid structural foundation that will support more accurate computational work and guide the design and interpretation of further experimental studies using CDO and small molecule CDO mimics.

Materials and Methods

Expression, purification and crystallization

Expression, purification and crystallization of recombinant rat CDO was done as described previously (17, 32, 46), but with seeds from crushed CDO crystals added to facilitate nucleation. The enzyme was stored at 8 mg/mL in 10 mM Tris pH=7.4 buffer and crystallized at room temperature in hanging drops formed from 0.5 μL seed stock in reservoir buffer,1 μL protein stock and 1 μL of a 0.1M tri-sodium citrate pH=5.6, 24% PEG 4000, and 0.15 M ammonium acetate reservoir (final pH=6.2) (17, 32). The crystals were isomorphous with those previously described (17, 46), having spacegroup $P4_32_12$ and one chain in the asymmetric unit.

Crystal soaks and data collection

Soaks were done at room temperature using the reservoir condition as an artificial mother liquor (AML). For cryo-crystallographic data collection, crystals were flash frozen by plunging into liquid nitrogen. For room temperature data collection, crystals were soaked as described and then scooped using a MicroRT™ capillary system (Mitegen).

For an initial pH series, crystals were incubated for 1 h in solutions with 100 mM cysteine added to AML that had its pH adjusted to 4.5, 5.5, 6.5, 7.5, 8.5 or 9.5. After data were collected, we discovered that the added Cys had altered the pH; by recreating the solutions, we found that the final pH values were 5.5, 6.2, 6.8, 7.0, 7.0 and 7.0. Additional 1 h Cys soaks were carried out in solutions made with final pH values (after cysteine addition) of 4.0, 4.5, 5.0, 7.0, 8.0, or 9.0. Crystals did not diffract after some 1 h soaks at pH=4.0 and 8.0 so shorter soaks of 15 or 20 min were done. Also, at pH=4.5 a series of short control soaks ranging from 30 s to 12 min were performed in addition to the 1 h soak. Ligand-free structures at a few pH values were

obtained from crystals soaked 40 – 60 min in AML at the targeted pH. For the ligand-free pH=8.0 structure, 100 mM Ala was added to AML to create buffering more equivalent to the Cys soak. For soaks at pH=8.0 with dithionite, crystals were incubated in a Cys- (20 min) or Ala- (60 min) containing degassed AML with 20 mM dithionite and 200 μM methyl viologen (as a color indicator of the redox status of the solution) and the final pH adjusted to pH=8 with degassed NaOH. Although crystals were soaked and handled in an aerobic environment, solutions remained deep blue for the duration of the soak.

Synchrotron data sets were collected at beamlines 5.0.1, 5.0.2, and 8.2.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) and beamline X26C at the National Synchrotron Light Source (Brookhaven National Laboratory). In-house diffraction data were collected on a Rigaku RU-H3R rotating Cu-anode operating at 50 kV and 100 mA and an R-Axis IV image plate detector. Data were processed and scaled using iMosflm(47) and SCALA.(48) For all refinements, the same 10% of the data as used previously(17) were flagged for use in R_{free} . Also, a random 10% of reflections beyond the previous 1.42 Å resolution limit were flagged for Rfree calculations. Data statistics are in Tables 3.1 and 3.2.

Crystallographic refinement

The unit cell constants for individual crystals were mostly within the ranges a=b=57.6-58.2 Å and c=121.7-122.9 Å, and these refinements all used a=b=57.60 Å and c=122.40 Å. Refinements were done using PHENIX(49) and for each dataset were initiated using an automated protocol that when applied to a refinement of the previously published 1.42 Å resolution CDO:Cys complex (PDB code 3eln) yielded a better R/R_{free}.(36) Refinements began with the 1.5 Å resolution unliganded CDO model (PDB code 2b5h) with, when appropriate, the coordinates for the Cys-persulfenate or cysteine added. The occupancy of the persulfenate was fixed at 0.7 in all persulfenate-containing models. The Cys-only models at pH=8 and 9 set the Cys occupancy at 1.0. Geometric restraints for the ligand as well as the Cys-Tyr crosslink were created by Elbow (50) and manual editing to match restraints used

previously.(17) For the OA-OB bond, the restraint of 1.47 Å was initially used (17), but was changed to 1.8 Å with a relatively lose weight (estimated standard deviation=0.10 Å) after observing a distinct density peak for the OB atom in some of the structures (eg. see Figure 3.2 and 3.5).

The first stage of refinement specified isotropic B-factors, riding hydrogens, automatic water adding, and rotamer fitting. This result was then further refined using anisotropic B-factors. The anisotropic B-factors were only accepted if their use resulted in an Rfree improvement of at least 1% compared with isotropic plus TLS refinements with 1 TLS group defining the entire chain. After these automated stages, the refinements were completed using manual rebuilding in Coot(51) and PHENIX-refine minimization. Some water sites were removed and others were added using standard criteria (>1 ρ_{rms} intensity in the $2F_o - F_c$ map, >2.4 Å distance from nearest contact), and all difference map peaks above 5 ρ_{rms} were checked. Molprobity(52) was used to find problems with model geometry in these final rounds. Refinement statistics are in Tables 3.1 and 3.2.

Estimates of coordinate uncertainties for individual atoms were calculated using the Cruickshank method (53) implemented in Refmac (54, 55). As examples of the coordinate uncertainties in the active site, for the pH=5.5+Cys / pH=9.0 Cys-only structures the residual uncertainties (to one significant figure or the nearest 0.01 Å and in order of the two structures) are as follows: Fe 0.01 / 0.005 Å; Tyr157-OH 0.04 / 0.03 Å; Arg60-NH2 0.16 / 0.09 Å; Cys-Sγ 0.04 / 0.02 Å; and for the persulfenate-OA and OB atoms in the pH=5.5 structure, 0.07 and 0.10 Å.

Acknowledgements.

We thank Dr. Dale Tronrud for useful scientific discussions and help with methods. We also thank Dr. Babak Andi and Dr. Allan Orville for collecting coordinated spectroscopic and diffraction data at the National Synchrotron Light Source (NSLS). This project was supported in part by Grant DK-056649 to MHS and PAK from the National Institute of Diabetes and Digestive and Kidney Diseases. It

was also aided by the Collaborative Crystallography Project of the Berkeley Center for Structural Biology supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. Synchrotron data were collected at the Advanced Light Source and the NSLS, respectively supported by contracts DE-AC02-98CH10886 and DE-AC02-05CH11231 from the Office of Basic Energy Sciences of the U.S. Department of Energy

Accession Numbers

Coordinates and structure factors for the CDO models have been deposited in the Protein Data Bank with accession numbers as follows: pH=4.0 + Cys 20 min (PDB code 4IEO); pH=4.5 + Cys 1 h (PDB code 4IEP); pH=5.0 + Cys 1 h (PDB code 4IEQ); pH=5.5 + Cys 1 h (PDB code 4IER); pH=6.2 + Cys 1 h (PDB code 4IES); pH=6.8 + Cys 1 h (PDB code 4IET); pH=7.0 + Cys 1 h (PDB code 4IEU); pH=8.0 + Cys 1 h (PDB code 4IEV); pH=9.0 + Cys 1 h (PDB code 4IEW); pH=6.2 + Lab-RT (PDB code 4IEX); pH=7.0 + Cys Lab-LT (PDB code 4IEY); pH=8.0 + Ala (PDB code 4IEZ); pH=8.0 + Ala + DT (PDB code 4JTN); pH=8.0 + Cys + DT (PDB code 4JTO).

Table 3.1. Data collection and Refinement Statistics for CDO pH series[a]

	pH=4.0 Cys	**pH=4.5 Cys**	**pH=5.0 Cys**	**pH=5.5 Cys**	**pH=6.2 Cys**	**pH=6.8 Cys**	**pH=7.0 Cys**	**pH=8.0 Cys**	**pH=9.0 Cys**
Data collection									
Resolution (Å)	34-1.55 (1.63-1.55)	30-1.45 (1.53-1.45)	34-1.40 (1.48-1.40)	42-1.45 (1.53-1.45)	29-1.40 (1.48-1.40)	42-1.40 (1.48-1.40)	42-1.25 (1.32-1.25)	42-1.60 (1.69-1.60)	34-1.45 (1.53-1.45)
Unique Obs.	28050 (4401)	33437 (4190)	34382 (3149)	33402 (2956)	41131 (5551)	39197 (5029)	53507 (6668)	28421 (4044)	33886 (3989)
Multiplicity	26.6 (24.8)	14.4 (6.5)	19.0 (8.6)	15.9 (6.4)	12.6 (4.7)	21.6 (6.7)	18.6 (7.0)	27.4 (27.9)	25.5 (15.4)
Completeness	91.0 (100)	88.9 (78.2)	82.3 (53.4)	88.3 (55.3)	99.1 (94.3)	93.0 (83.4)	92.1 (80.0)	100 (100)	88.2 (73.3)
$\langle I/\sigma \rangle$	12.9 (0.5)	10.3 (0.9)	12.0 (0.9)	18.3 (0.5)	17.5 (0.6)	13.8 (0.6)	19.2 (0.6)	11.4 (0.4)	16.4 (0.5)
R_{meas}[b] (%)	17.4 (812)	12.9 (303)	13.2 (406)	8.0 (428)	7.9 (294)	11.5 (277)	9.9 (365)	24.9(1109)	12.7 (654)
R_{pim} (%)	3.4 (162)	3.3 (117)	3.0 (137)	1.9(162)	2.1 (129)	2.1 (105)	1.9 (135)	4.7 (208)	2.5 (161)
$CC_{1/2}$[c] (%)	1.00 (0.12)	1.00 (0.13)	1.00 (0.13)	1.00(0.13)	1.00 (0.15)	1.00 (0.23)	1.00 (0.15)	1.00 (0.12)	1.00 (0.10)
Res $\langle I/\sigma \rangle \sim 2$ (Å)[d]	1.80	1.60	1.50	1.60	1.55	1.60	1.40	1.85	1.60
Refinement									
R_{cryst} / R_{free} (%)	20.3/24.1	15.2/20.6	14.3/19.4	13.7/19.8	14.5/18.5	15.3/20.1	14.5/17.6	18.0/21.8	14.9/19.9
No. Obs	27242	32582	33980	32809	40816	38127	52891	27709	32890
No. residues	186	186	186	186	186	186	186	186	186
No. waters	132	165	216	210	240	214	283	190	258
No. atoms	1723	1814	1837	1761	1808	1789	1874	1721	1802
rmsd angles (°)	1.44	1.23	1.26	1.25	1.23	1.26	1.28	1.33	1.30
rmsd lengths (Å)	0.013	0.011	0.011	0.013	0.012	0.012	0.011	0.013	0.014
φ,ψ-favored (%)[e]	98.5	99.5	98.9	99.5	99.5	97.9	98.9	98.9	99.5
φ,ψ-outliers (%)[e]	0	0	0	0	0	0	0	0	0
$\langle B \rangle$ protein (Å^2)	27	23	22	29	21	25	22	24	25
$\langle B \rangle$ ligand (Å^2)	-	-	-	33	18	21	18	30	20
PDB code	4IEO	4IEP	4IEQ	4IER	4IES	4IET	4IEU	4IEV	4IEW

[a] All refinements used space group $P4_32_12$ with a=b=57.60 Å and c= 122.40 Å. Numbers in parentheses refer to the highest resolution bin.
[b] R_{meas} is the multiplicity-weighted merging R-factor[55]
[c] $CC_{1/2}$ is the correlation between two datasets each based on half of the data as defined in Karplus & Diederichs.[35]
[d] The resolution at which $\langle I/\sigma \rangle \sim 2$ is given to allow comparison with previous standards for selecting high resolution limits.
[e]Ramachandran statistics as defined by Molprobity[51]

Table 3.2. Data collection and Refinement Statistics for CDO additional datasets[a]

	pH=6.2 Lab-RT	**pH=7.0 Cys Lab-LT**	**pH=8.0 Ala 40 min**	**pH=8.0 Cys + DT 15 min**
Data collection				
Resolution (Å)	29-2.15 (2.27-2.15)	33-1.63 (1.72-1.63)	31-1.40 (1.48-1.40)	29-2.00 (2.11-2.00)
Unique Obs.	11697 (1594)	26578 (3553)	37124 (4475)	13313 (1708)
Multiplicity	2.5 (2.5)	20.9 (13.9)	11.4 (4.5)	7.6 (8.0)
Completeness	96.1 (92.0)	98.6 (92.9)	87.3 (74.1)	90.0 (82.4)
$\langle I/\sigma \rangle$	3.6 (0.5)	8.2 (0.7)	9.6 (0.5)	11.2 (2.7)
R_{meas}[b] (%)	27.1 (252)	23.9 (662)	12.0 (388)	11.6 (67.2)
R_{pim} (%)	16.1 (149)	5.1 (175)	3.4 (167)	4.2 (23.1)
$CC_{1/2}$[c] (%)	0.98 (0.13)	1.00 (0.11)	1.00 (0.14)	1.00 (0.85)
Res $\langle I/\sigma \rangle \sim 2$ (Å)[d]	2.80	1.85	1.60	-
Refinement				
R_{cryst} / R_{free} (%)	17.5/25.6	21.3/25.6	15.3/20.8	17.6/23.3
No. Obs.	11021	25288	32519	12831
No. residues	186	186	186	186
No. waters	53	247	229	127
No. atoms	1558	1788	1778	1644
rmsd angles (°)	1.59	1.50	1.28	1.37
rmsd lengths (Å)	0.017	0.015	0.014	0.014
φ,ψ-favored (%)[e]	98.91	97.3	99.0	99.5
φ,ψ-outliers (%)[e]	0	0	0	0
$\langle B \rangle$ protein (Å^2)	41	30	24	33
$\langle B \rangle$ ligand (Å^2)	-	27	-	32
PDB code	4IEX	4IEY	4IEZ	4JTO

[a] All refinements used space group $P4_32_12$, with cell constants a=b=57.60 Å and c= 122.40 Å. Numbers in parentheses refer to the highest resolution bin.
[b] R_{meas} is the multiplicity-weighted merging *R*-factor[55]
[c]$CC_{1/2}$ is the correlation between two datasets each based on half of the data as defined in Karplus & Diederichs.[35]
[d] The resolution at which $\langle I/\sigma \rangle \sim 2$ is given to allow comparison with previous standards for selecting high resolution limits.
[e]Ramachandran statistics as defined by Molprobity[51]

Table 3.3. Comparing standard and $CC_{1/2}$-based resolution cutoffs for pH=7.0 + Cys data[a].

	R_{meas}-based	**I/σ-based**	**$CC_{1/2}$-based**
Data collection			
Resolution (Å)[b]	1.58-1.50	1.48-1.40	1.32-1.25
$CC_{1/2}$	0.92	0.71	0.15
Average I/σ	3.9	1.9	0.6
R_{meas}[c]	61%	119%	365%
R/R_{free} @1.5 Å (%)[d]	14.5/17.6	14.7/17.6	15.3/17.6 (14.5/17.6)[e]
Waters built	237	253	251 (283)

[a] The pH 7.0 Cys-bound models were refined using the same automatic phenix.refine strategy which was used as the first step in refinement (see methods) against data truncated at 1.50 Å (based on R_{meas} ~60%), 1.40 Å (based on I/σ ~2), or 1.25 Å (based on $CC_{1/2}$ criteria)[35].
[b] The resolution range for the data used in refinement was 42 Å to the high resolution limit
[c] the multiplicity-weighted merging *R*-factor[55]
[d] R_{cryst} and R_{free} of the indicated model calculated at 1.50 Å resolution
[e] Numbers in parentheses are for the final refined model

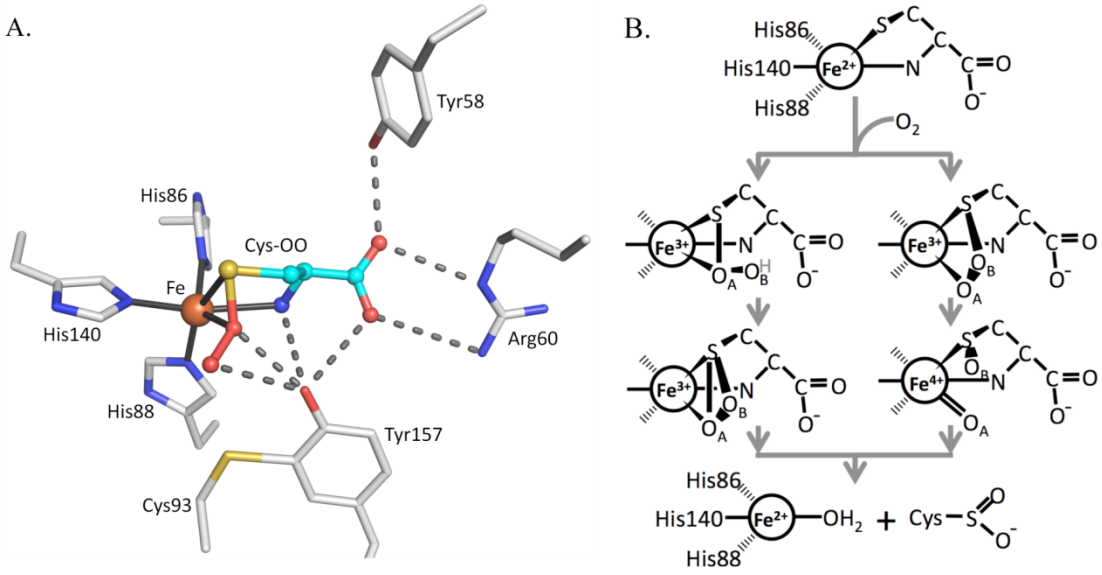


Figure 3.1. CDO active site geometry and two proposed mechanisms. (A) The CDO active site residues (white carbons) with bound Cys-persulfenate (teal carbons) (PDB entry 3ELN) showing potential H-bond interactions (dashes) and the coordination bonds (black line) to Fe (orange). (B) The key distinction between the two types of mechanisms being considered for CDO is shown. After dioxygen binds to Fe(II)-Cys CDO (top image), one route (left side) involves persulfenate/persulfenic acid formation, where the iron-proximal oxygen (OA) attacks the Cys-Sγ atom.(17) The H-atom shown bound to OB (grey text) may or may not be present in this proposed intermediate. The other route (right side), similar to monooxygenase chemistry, proceeds by attack of the Fe-distal oxygen (OB) on the Cys-Sγ atom, followed by creation of a reactive Fe(IV)-oxo intermediate as shown.(20, 22)

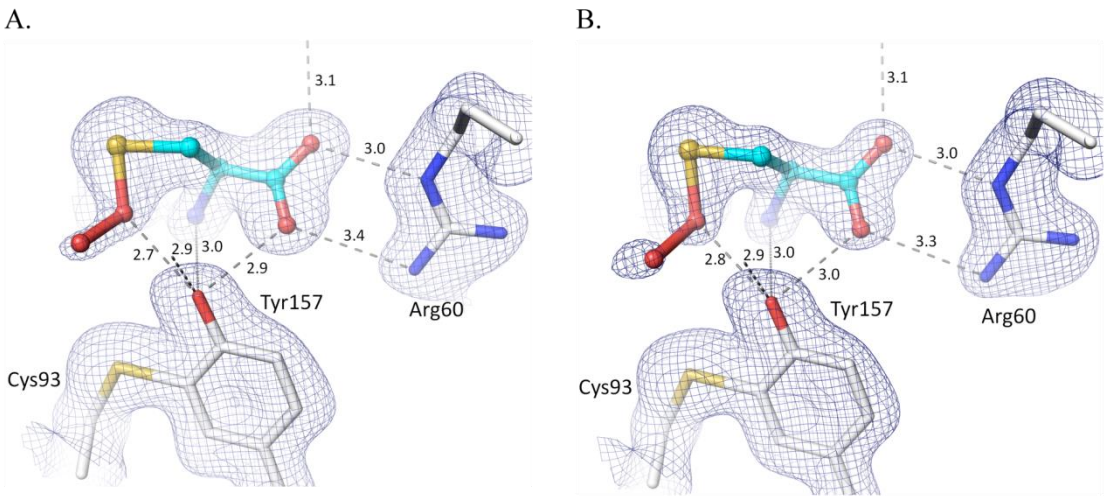


Figure 3.2. Electron density map improvement using a more generous high resolution cutoff. (A) Shown is $2F_o$-F_c active site electron density for the pH=6.8+Cys soak at 1.60 Å resolution (contoured at 1.0 ρ_{rms}) along with the refined model (atom coloring as in Figure 3.1). At this resolution, the $\langle I/\sigma \rangle$ is 2.1, R_{meas} is 154%, and $CC_{1/2}$ is 0.78. (B) Same as (A) but using data out to 1.40 Å resolution with the map contoured at 1.2 ρ_{rms} (see statistics in Table 3.1). Many details of the map show subtle improvement, but especially notable is the improvement for the distal persulfenate oxygen, which becomes resolved from the rest of the persulfenate.

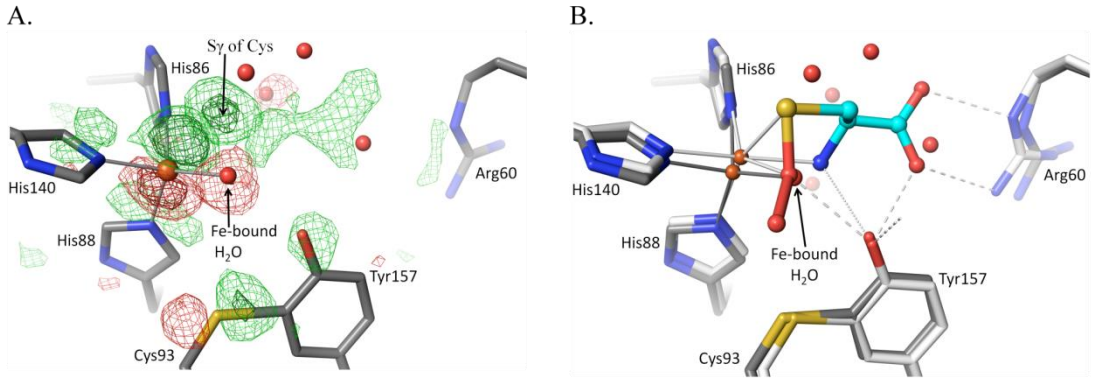


Figure 3.3. Active site changes upon Cys binding. (A) F_o-F_o difference map between the pH 6.2-Cys soak and unliganded CDO (PDB entry 2B5H; Simmons 2006) using phases calculated from the unliganded structure. The unliganded model is shown (grey carbons) and the map contour levels are ±4.0 ρ_{rms} (light green/red) and ±10.0 ρ_{rms} (dark green/red). (B) Overlay, in the same view as panel A, of the models of unliganded (grey carbon atoms) and Cys-persulfenate bound (white/cyan carbon atoms) CDO. All waters shown are from the unliganded structure and are displaced by the bound ligand. The visible shifts of the iron and other active site residues and the placement of the persulfenate are consistent with the difference density shown in panel A.

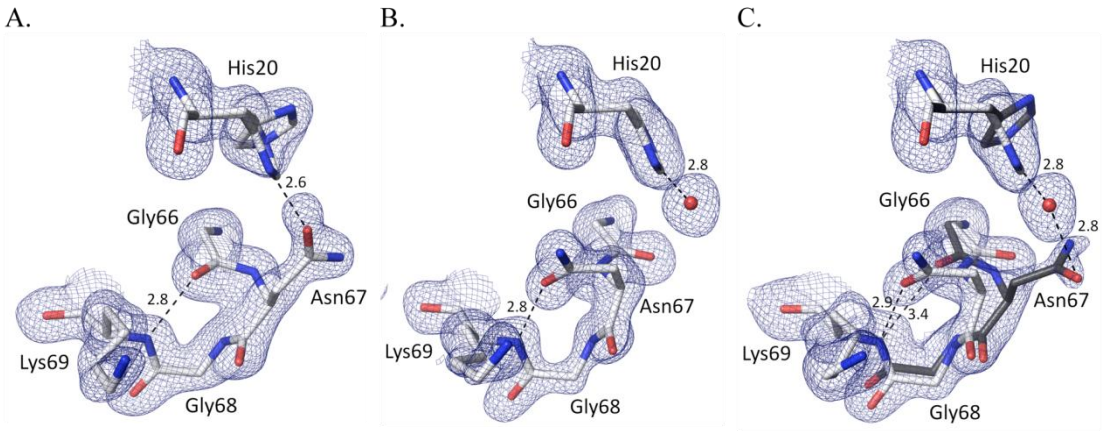


Figure 3.4. pH-dependent changes in the Asn67 loop. (A) $2F_o$-F_c electron density and refined model from the pH5.5-Cys soak for residues 66 – 69 and His20; contoured at 1.2 ρ_{rms}, with potential H-bond interactions (dashed lines; distance in Å) shown. (B) Equivalent image based on the pH9-Cys soak. (C) Equivalent image based on the pH7-Cys soak contoured at 0.8 ρ_{rms}. The estimated occupancies of His20 alternate conformations are 100%/0% at pH=9.0, 80%/20% at pH=7.0, and a 50%/50% at pH=5.5. Assuming a protonated His20 adopts a 50/50 mix of conformations, the occupancies of the two turn conformations imply that His20 would be ~0, ~60 and ~100% deprotonated at pH=5.5, 7.0 and 9.0, respectively. Also at pH≤4.5 (not shown) His20 adopts a single conformation, a change we attribute to a protonation of the Asp64 carboxylate, which H-bonds to His20-Nε2 in all structures.

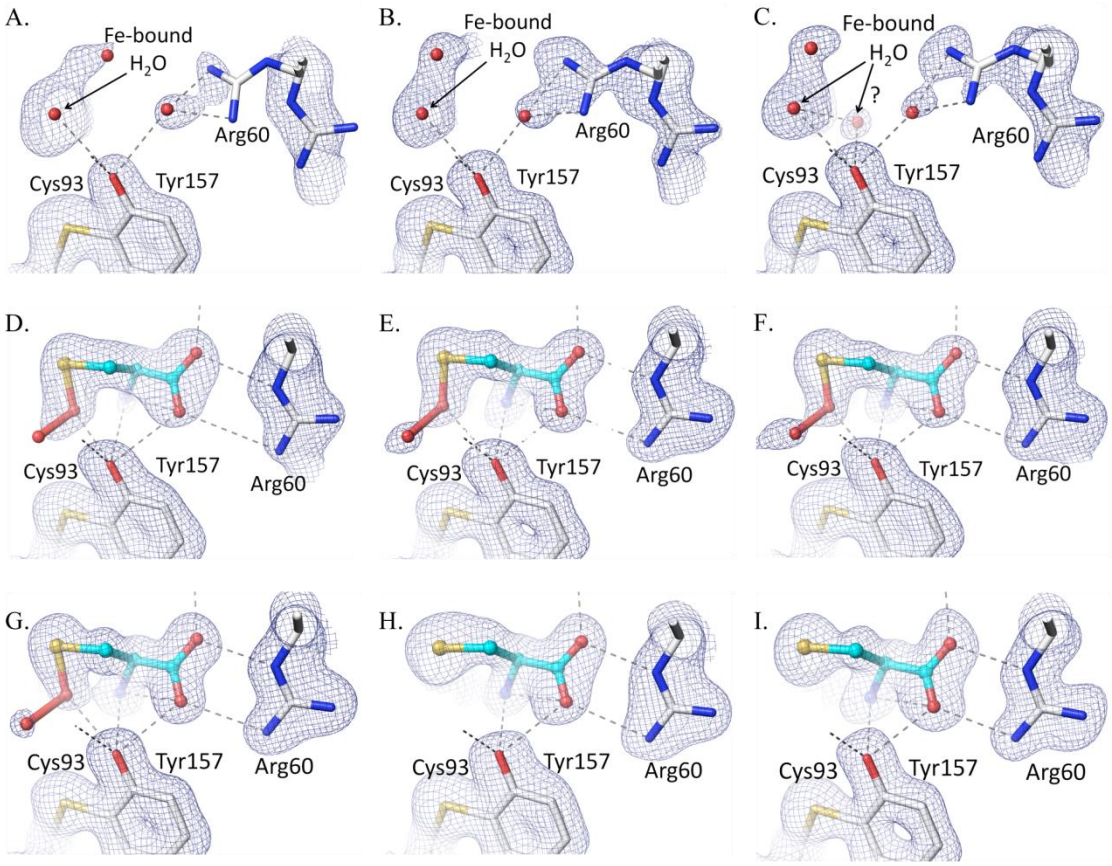


Figure 3.5. Active site density of Cys soaks as a function of pH. Each panel shows, for the relevant structure, the refined model (white carbon atoms for protein and cyan carbons for the ligand, and potential H-bond interactions as dashed lines) and the $2F_o$-F_c density contoured at 1.2 ρ_{rms}. Shown are Cys soaks at (A) pH=4.0, (B) pH=4.5, (C) pH=5.0, (D) pH=5.5, (E) pH=6.2, (F) pH=6.8, (G) pH=7.0, (H) pH=8.0, and (I) pH=9.0. Although His155-Nε2 is not shown, the initial trajectory of its H-bond to the Tyr157 hydroxyl is also shown.

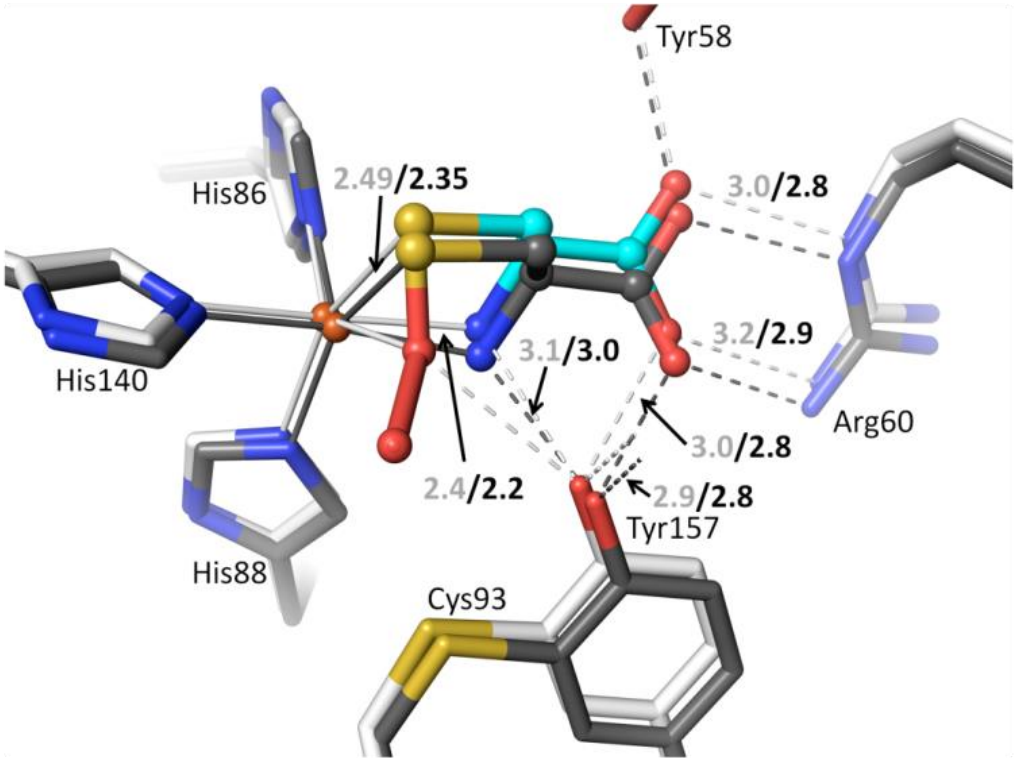


Figure 3.6. Active site differences between Cys-bound and Cys-persulfenate bound CDO. The persulfenate complex at pH=6.2 (cyan and white carbons) is overlaid on the Cys-bound complex pH=9.0 (dark carbons). Coordination and H-bond distances are given for the persulfenate complex (light text) and for the Cys-bound complex (black text). In the Cys-bound complex the Cys N- and Sγ-ligands are moved slightly to partly occupy the space left by the missing sixth ligand. This movement causes a slight shift in the whole Cys orientation and of the Cys-Tyr and Arg that interact with it.

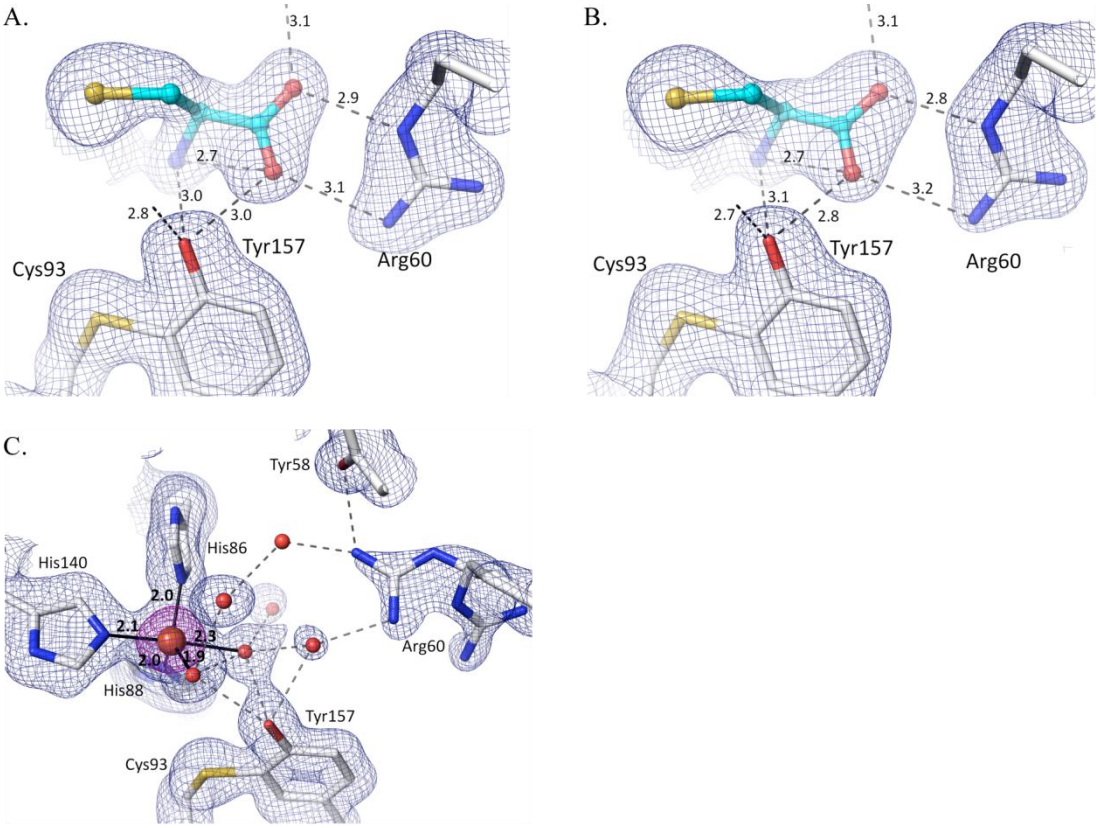


Figure 3.7. Cys-only complex and 5-coordinate iron at pH=8.0 with and without dithionite. (A) Structure and $2F_o$-F_c density for the pH=8.0 Cys soak contoured at 1.0 ρ_{rms} (same as figure 3.5H but with H-bond distances given). (B) Same as A, but for the pH=8.0 + Cys + dithionite soak. (C) $2F_o$-F_c density contoured at 1.4 ρ_{rms} (blue) and 5.0 ρ_{rms} (purple) for ligand-free CDO at pH=8.0. Iron ligating bonds (solid black lines) are shown with distances.

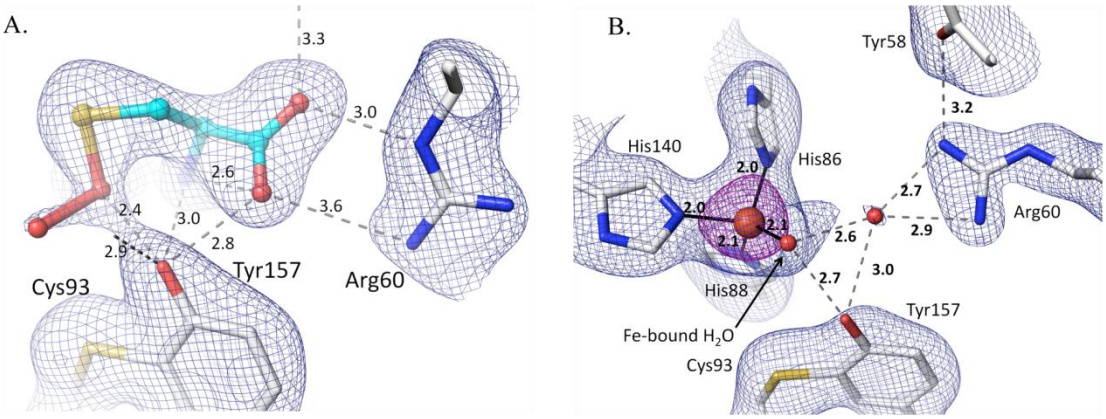


Figure 3.8. The persulfenate is not an artifact of freezing or synchrotron radiation. Each panel shows $2F_o$-F_c density along with the corresponding refined model with protein carbons (white) and ligand carbons (cyan) and H-bonds (dashed lines with distances). (A) Lab-LT-Cys structure with density for bound Cys-persulfenate contoured at 0.7 ρ_{rms}. (B) Lab-RT structure of unliganded CDO with density contoured at 1.4 ρ_{rms} (blue) and 5.0 ρ_{rms} (purple). The view in panel B matches that of figure 3.6C and D.

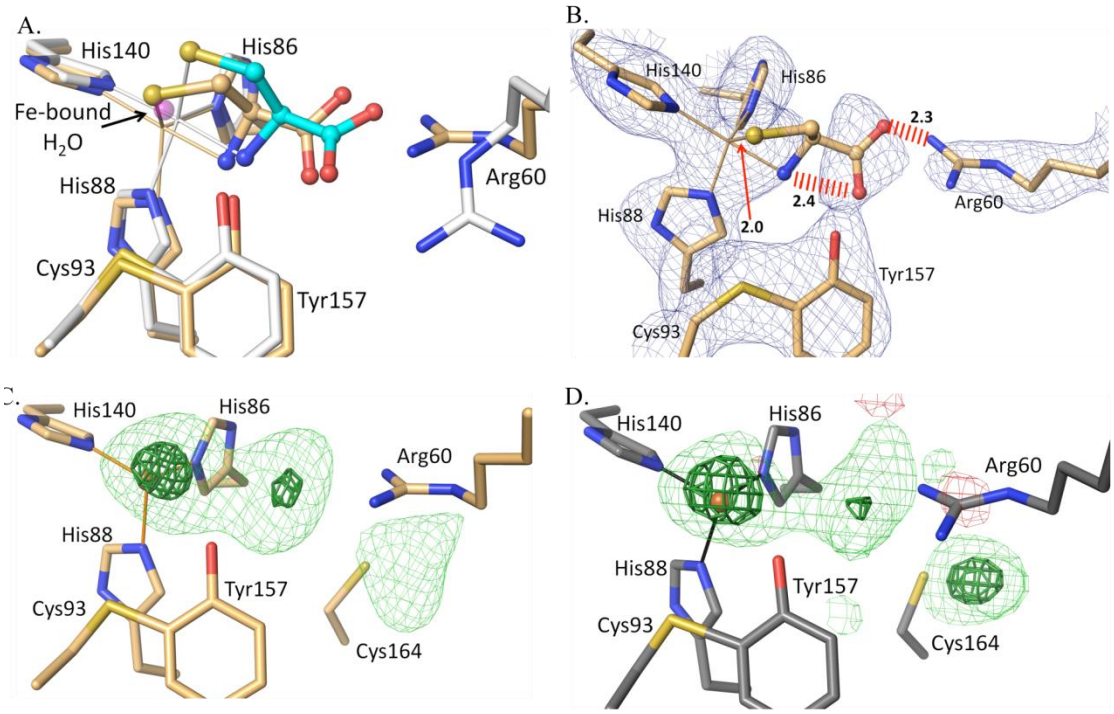


Figure 3.9. Reevaluation of the published human CDO-Cys complex. (A) An overlay of active site structure of PDB entry 2IC1 (orange carbon atoms and orange bonds to the iron atom) on the pH=9 CDO-Cys complex reported here (white and cyan carbon atoms and white bonds to the iron atom), with the iron bound water from unliganded CDO (PDB entry 2GH2) shown for reference (semi-transparent purple atom) and with the iron positions indicated by the apex of the coordination bonds. (B) A $2F_o$-F_c omit map of the active site of 2IC1 (calculated after omitting the ligand from the model) contoured at 1.5 ρ_{rms}. Also indicated are three unreasonably close approaches in the 2IC1 structure (red hash marks or arrow with distances). Note that at this contour level density for the Cys in 2IC1 is not continuous even though the iron ligating His residues and other active site residues have strong continuous density. (C) Difference map for PDB entry 2IC1 (Ye et al. 2007) with the ligand atoms removed from the model; the contour levels are 4.0 ρ_{rms} (lightgreen) and 10.0 ρ_{rms} (darkgreen). (D) Same as C, but calculated for PDB entry 2GH2 (unliganded disulfide modified rat CDO)(10) after removing active site waters 505, 651, 656, 828, 830 and truncating the data to 2.7 Å resolution to match that of PDB entry 2IC1.

References

1. Stipanuk, M. H., Simmons, C. R., Karplus, P. A., and Dominy, J. E., Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. Amino Acids 41, 91-102.

2. Bagley, P. J., and Stipanuk, M. H. (1994) The activities of rat hepatic cysteine dioxygenase and cysteinesulfinate decarboxylase are regulated in a reciprocal manner in response to dietary casein level. J Nutr 124, 2410-2421.

3. Dominy, J. E., Jr., Hwang, J., and Stipanuk, M. H. (2007) Overexpression of cysteine dioxygenase reduces intracellular cysteine and glutathione pools in HepG2/C3A cells. Am J Physiol Endocrinol Metab 293, E62-69.

4. Dominy, J. E., Jr., Hwang, J., Guo, S., Hirschberger, L. L., Zhang, S., and Stipanuk, M. H. (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. J Biol Chem 283, 12188-12201.

5. Stipanuk, M. H. (2004) Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annu Rev Nutr 24, 539-577.

6. Ueki, I., Roman, H. B., Valli, A., Fieselmann, K., Lam, J., Peters, R., Hirschberger, L. L., and Stipanuk, M. H. (2011) Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize taurine and an increased catabolism of cysteine to hydrogen sulfide. Am J Physiol Endocrinol Metab 301, E668-684.

7. Heafield, M. T., Fearn, S., Steventon, G. B., Waring, R. H., Williams, A. C., and Sturman, S. G. (1990) Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. Neurosci Lett 110, 216-220.

8. Bradley, H., Gough, A., Sokhi, R. S., Hassell, A., Waring, R., and Emery, P. (1994) Sulfate metabolism is abnormal in patients with rheumatoid arthritis. Confirmation by in vivo biochemical findings. J Rheumatol 21, 1192-1196.

9. Brait, M., Ling, S., Nagpal, J. K., Chang, X., Park, H. L., Lee, J., Okamura, J., Yamashita, K., Sidransky, D., and Kim, M. S. (2012) Cysteine dioxygenase 1 is a tumor suppressor gene silenced by promoter methylation in multiple human cancers. PLoS One 7, e44951.

10. Simmons, C. R., Liu, Q., Huang, Q., Hao, Q., Begley, T. P., Karplus, P. A., and Stipanuk, M. H. (2006) Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. J Biol Chem 281, 18723-18733.

11. McCoy, J. G., Bailey, L. J., Bitto, E., Bingman, C. A., Aceti, D. J., Fox, B. G., and Phillips, G. N., Jr. (2006) Structure and mechanism of mouse cysteine dioxygenase. Proc Natl Acad Sci U S A 103, 3084-3089.

12. Kleffmann, T., Jongkees, S. A., Fairweather, G., Wilbanks, S. M., and Jameson, G. N. (2009) Mass-spectrometric characterization of two posttranslational modifications of cysteine dioxygenase. J Biol Inorg Chem 14, 913-921.

13. Dominy, J. E., Jr., Simmons, C. R., Karplus, P. A., Gehring, A. M., and Stipanuk, M. H. (2006) Identification and characterization of bacterial cysteine dioxygenases: a new route of cysteine degradation for eubacteria. J Bacteriol 188, 5561-5569.

14. Siakkou, E., Rutledge, M. T., Wilbanks, S. M., and Jameson, G. N. (2011) Capturing crosslink formation with enzymatic activity in cysteine dioxygenase. Biochim Biophys Acta.

15. Imsand, E. M., Njeri, C. W., and Ellis, H. R. (2012) Addition of an external electron donor to in vitro assays of cysteine dioxygenase precludes the need for exogenous iron. Arch Biochem Biophys 521, 10-17.

16. Ye, S., Wu, X., Wei, L., Tang, D., Sun, P., Bartlam, M., and Rao, Z. (2007) An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. J Biol Chem 282, 3391-3402.

17. Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Jr., Begley, T. P., Stipanuk, M. H., and Karplus, P. A. (2008) A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. Biochemistry 47, 11390-11392.

18. Pierce, B. S., Gardner, J. D., Bailey, L. J., Brunold, T. C., and Fox, B. G. (2007) Characterization of the nitrosyl adduct of substrate-bound mouse cysteine dioxygenase by electron paramagnetic resonance: electronic structure of the active site and mechanistic implications. Biochemistry 46, 8569-8578.

19. Crawford, J. A., Li, W., and Pierce, B. S. (2011) Single turnover of substrate-bound ferric cysteine dioxygenase with superoxide anion: enzymatic reactivation, product formation, and a transient intermediate. Biochemistry 50, 10241-10253.

20. Gardner, J. D., Pierce, B. S., Fox, B. G., and Brunold, T. C. (2010) Spectroscopic and computational characterization of substrate-bound mouse cysteine dioxygenase: nature of the ferrous and ferric cysteine adducts and mechanistic implications. Biochemistry 49, 6033-6041.

21. Aluri, S., and de Visser, S. P. (2007) The mechanism of cysteine oxygenation by cysteine dioxygenase enzymes. J Am Chem Soc 129, 14846-14847.

22. Kumar, D., Thiel, W., and de Visser, S. P. (2011) Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. J Am Chem Soc 133, 3869-3882.

23. de Visser, S. P., and Straganz, G. D. (2009) Why do cysteine dioxygenase enzymes contain a 3-His ligand motif rather than a 2His/1Asp motif like most nonheme dioxygenases? J Phys Chem A 113, 1835-1846.

24. Mirza, S. A., Pressler, M. A., Kumar, M., Day, R. O., and Maroney, M. J. (1993) Oxidation of Nickel Thiolate Ligands by Dioxygen. Inorganic Chemistry 32, 977-987.

25. Clennan, E. L. (2001) Persulfoxide: Key intermediate in reactions of singlet oxygen with sulfides. Accounts Chem Res 34, 875-884.

26. Grapperhaus, C. A., and Darensbourg, M. Y. (1998) Oxygen capture by sulfur in nickel thiolates. Accounts Chem Res 31, 451-459.

27. Farmer, P. J., Solouki, T., Soma, T., Russell, D. H., and Darensbourg, M. Y. (1993) Divergent Pathways for the Addition of Dioxygen to Sulfur in Nickel Cis-Dithiolates - an Isotopomeric Analysis. Inorganic Chemistry 32, 4171-4173.

28. Kumar, D., Sastry, G. N., Goldberg, D. P., and de Visser, S. P. (2012) Mechanism of S-oxygenation by a cysteine dioxygenase model complex. J Phys Chem A 116, 582-591.

29. McQuilken, A. C., Jiang, Y., Siegler, M. A., and Goldberg, D. P. (2012) Addition of dioxygen to an N4S(thiolate) iron(II) cysteine dioxygenase model gives a structurally characterized sulfinato-iron(II) complex. J Am Chem Soc 134, 8758-8761.

30. McQuilken, A. C., and Goldberg, D. P. (2012) Sulfur oxygenation in biomimetic non-heme iron-thiolate complexes. Dalton Trans 41, 10883-10899.

31. Chai, S. C., Jerkins, A. A., Banik, J. J., Shalev, I., Pinkham, J. L., Uden, P. C., and Maroney, M. J. (2005) Heterologous expression, purification, and characterization of recombinant rat cysteine dioxygenase. J Biol Chem 280, 9865-9869.

32. Simmons, C. R., Hirschberger, L. L., Machi, M. S., and Stipanuk, M. H. (2006) Expression, purification, and kinetic characterization of recombinant rat cysteine dioxygenase, a non-heme metalloenzyme necessary for regulation of cellular cysteine levels. Protein Expr Purif 47, 74-81.

33. Chai, S. C., Bruyere, J. R., and Maroney, M. J. (2006) Probes of the catalytic site of cysteine dioxygenase. J Biol Chem 281, 15774-15779.

34. Siakkou, E., Wilbanks, S. M., and Jameson, G. N. (2010) Simplified cysteine dioxygenase activity assay allows simultaneous quantitation of both substrate and product. Anal Biochem 405, 127-131.

35. Tchesnokov, E. P., Wilbanks, S. M., and Jameson, G. N. (2012) A strongly bound high-spin iron(II) coordinates cysteine and homocysteine in cysteine dioxygenase. Biochemistry 51, 257-264.

36. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. Science 336, 1030-1033.

37. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008) Data mining of metal ion environments present in protein structures. J Inorg Biochem 102, 1765-1776.

38. Morgan, B., and Lahav, O. (2007) The effect of pH on the kinetics of spontaneous Fe(II) oxidation by O_2 in aqueous solution--basic principles and a simple heuristic description. Chemosphere 68, 2080-2084.

39. Green, M. T. (2006) Application of Badger's rule to heme and non-heme iron-oxygen bonds: an examination of ferryl protonation states. J Am Chem Soc 128, 1902-1906.

40. Weik, M., Berges, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houee-Levin, C., and Ravelli, R. B. (2002) Evidence for the formation

of disulfide radicals in protein crystals upon X-ray irradiation. J Synchrotron Radiat 9, 342-346.

41. Hall, A., Sankaran, B., Poole, L. B., and Karplus, P. A. (2009) Structural changes common to catalysis in the Tpx peroxiredoxin subfamily. J Mol Biol 393, 867-881.

42. Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., and Alber, T. (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc Natl Acad Sci U S A 108, 16247-16252.

43. Halle, B. (2004) Biomolecular cryocrystallography: structural changes during flash-cooling. Proc Natl Acad Sci U S A 101, 4793-4798.

44. Ibarra, C. A., Chowdhury, P., Petrich, J. W., and Atkins, W. M. (2003) The anomalous pKa of Tyr-9 in glutathione S-transferase A1-1 catalyzes product release. J Biol Chem 278, 19257-19265.

45. Schlegel, B. P., Jez, J. M., and Penning, T. M. (1998) Mutagenesis of 3 alpha-hydroxysteroid dehydrogenase reveals a "push-pull" mechanism for proton transfer in aldo-keto reductases. Biochemistry 37, 3538-3548.

46. Simmons, C. R., Hao, Q., and Stipanuk, M. H. (2005) Preparation, crystallization and X-ray diffraction analysis to 1.5 A resolution of rat cysteine dioxygenase, a mononuclear iron enzyme responsible for cysteine thiol oxidation. Acta Crystallogr Sect F Struct Biol Cryst Commun 61, 1013-1016.

47. Leslie, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EAMCB newsletter on protein crystallography 26.

48. Evans, P. R. (1997) Recent advances in phasing. In Proc. CCP4 Study Weekend (Wilson, K. S., Davies, G., Ashton, A. W., and Bailey, S., Eds.), pp 97–102, Daresbury Laboratory, Warrington, UK.

49. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213-221.

50. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr 65, 1074-1080.

51. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D 66, 486-501.

52. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol 285, 1711-1733.

53. Cruickshank, D. W. J. (1996) Protein precision re-examined: Luzzati plots do not estimate final errors. Proceedings of the CCP4 Study Weekend January 1996.

54. Murshudov, G. N., and Dodson, E. J. (1997) Simplified error estimation a la Cruickshank in macromolecular crystallography. CCP4 Newsletter January 1997.

55. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D 53, 240-255.

Chapter 4

Crystal Structures of Non-Crosslinked Variants of Mammalian Cysteine Dioxygenase with and without ligands

Camden M. Driggers, Lawrence L. Hirschberger, Martha H. Stipanuk,
and P. Andrew Karplus

In preparation for *Biochemistry*

Abstract

In mammals, high levels of cysteine (Cys) are toxic and are associated with neurodegenerative disorders and cysteine dioxygenase (CDO) is a non-heme iron enzyme that helps regulate Cys levels through converting Cys to Cys sulfinic acid (CSA). In addition to CDO being regulated by the amount of enzyme present, its activity is regulated through the formation of a Cys93-Tyr157 crosslink that increases its catalytic efficiency. To investigate how the crosslink influences activity, we have solved the crystal structures of C93A and Y157F mutants that do not form crosslinks. Surprisingly, we find that in the absence of Cys, a chloride ion binds to the active site iron in both the C93A and Y157F variants, and treatment of these crystals with dithionite results does not lead to loss of the chloride, although some is displaced by the dithionite decomposition product thiosulfate. Upon exposure of CDO crystals to Cys, Cys displaces the chloride but does not bind the same as it does to wild-type CDO: the Cys similarly coordinates the iron though its thiolate, but the α-amino group does not coordinate the iron. This defines a key role for Tyr157 in Cys binding, through both positioning Cys in the active site and modulating the pKa of its α-amino group. When the crosslink is disrupted, Tyr157 shifts and is not well positioned to participate in the catalytic triad. Cys-persulfenate does not form in the C93A and Y157F active sites, indicating the crosslink is necessary for persulfenate formation in the crystalline enzyme. We also investigated the structural basis for inhibition by homocysteine and azide. We found that homocysteine binds to the active sites of wild-type CDO and the variant, blocking the cysteine binding site and also blocking the oxygen binding site inferred from the wild-type persulfenate structures, but that azide only binds to the wild-type CDO, associating with the hydroxyl of Tyr157 in the crosslinked enzyme.

Introduction

Cysteine dioxygenase (CDO) is a mononuclear non-heme iron protein that catalyses the conversion of cysteine (Cys) to cysteine sulfinic acid (CSA) (Scheme 4.1) by incorporating both oxygen atoms of molecular oxygen to form CSA (reviewed in 1). That the dioxygen binding site is only formed after Cys binds was elucidated by showing that the oxygen surrogates NO (2) and CN (3) could only bind after Cys bound (2,3). Also, even though Fe(III)-CDO is not enzymatically active, spectroscopic titrations have shown that Fe(III)-CDO can bind Cys between pH=5.5 and 9.5 (4,5), and after Cys forms the catalytically inactive Fe(III)-Cys complex, superoxide can rescue activity although product formation is at a much slower rate than is reported for Fe(II) CDO (5). In mammals studied, the reaction catalyzed by CDO is the first and committed step of cysteine catabolism, and in response to cysteine levels the amount of CDO present varies over ~20-fold, primarily controlled by the ubiquitin-proteasome system (6,7).

Maintaining proper cysteine levels in the cell is important as Cys needs to be available (i.e. for protein and glutathione synthesis), but at high levels Cys is toxic by two reported mechanisms, the first by increasing oxidative damage (8), and the second by acting as an agonist of the NMDA-subtype of glutatmate receptor and destroying neurons in several areas of the central nervous system (9,10). Loss-of-function mutations in CDO have been linked to rheumatoid arthritis and some other neurodegenerative disorders based on observations of high cysteine and low sulfate levels (11-13). Also, high levels of homocysteine together with elevated amounts of cysteine have been observed in Alzheimer's disease (14), and this could be a result of the inhibition of CDO by homocysteine, which is a known inhibitor of CDO (as is azide (15)) (16).

In the $CDO^{-/-}$ mouse, cysteine and sulfate levels are in the normal range, but concentrations of taurine are low, and metabolism of cysteine through alternative desulfhydration pathways is increased, giving rise to evidence of H_2S toxicity (17). These $CDO^{-/-}$ mice have connective tissue abnormalities including enlarged alveolar

air spaces and abnormal organization of elastic fibers in the vasculature and parenchyma of lungs, high postnatal mortality, joint hyperlaxity and fatty acid oxidation defects (17). CDO has been implicated as a tumor suppressor as decreased expression of CDO due to promoter methylation was seen in multiple tumor tissues, and the forced expression of CDO in cancer cells slowed their growth (18). Recently, CDO was implicated as a marker of liposarcoma progression and adipogenic differentiation as well-differentiated liposarcomas from tissues were found to have higher levels of CDO mRNA and protein levels than de-differentiated liposarcomas and in cultures of human mesenchymal stem cells, CDO expression levels were up-regulated as cells underwent differentiation into mature adipocytes (19).

Initial structural studies of CDO showed it to have a cupin fold (20,21), with the structure of rat CDO having an unusual tetrahedral non-heme Fe coordination involving three (neutral (4)) His residues and a water as ligands (20). In contrast to this, a structure with Ni in the place of Fe has an octahedral geometry with three waters bound (21), and some have suggested that the octrahedral coordination might also be relevant for iron (4,15,22). CDO's three-His Fe coordination is unusual among mono-nuclear non-heme iron enzymes and has been suggested to be optimal for dioxygenation of substrate (23). Typically, non-heme iron enzymes in the cupin family have two His residues and one carboxylate residue (a Glu in the place of Cys93 of rat CDO) coordinating the iron on one face, with substrate or waters or cofactors binding on the opposite face. Another exception to this standard iron 2-His 1-carboxyl coordination is found in the aliphatic halogenase enzyme SyrB2, where a single chloride anion (Cl^-) replaces the carboxylate ligand; the chloride is not an unchanging part of the active site, but is actually a substrate and during catalysis is transferred to the methyl group of amino acids covalently bound to peptidyl carrier proteins (24,25).

Early crystal structures of mammalian CDO revealed the presence of a Cys93-Tyr157 thioether linkage (14; 15), which was later confirmed to be present in solution (26,27). This crosslink is not present in CDO from prokaryotes, which have a Gly at the Cys93 position (1;7; 22) (see also PDB entry 3eqe). CDO is present in both the

crosslinked and non-crosslinked form in cells and tissues (27); native or recombinant non-crosslinked CDO can be converted to the crosslinked form by incubation with ferrous iron, oxygen and cysteine (4,27,28). The autocatalyzed production the Cys93-Tyr157 crosslink is reported to increase activity ~10-fold (27) (or more (28,29)) and it has been suggested that this represents a purposeful posttranslational regulation strategy (27). If this change occurred together with a 20-fold variation in CDO levels (6,7), it could lead to an increase of ~200-fold (or more (28,29)) in CDO activity in response to high cysteine levels and play a key role in intracellular sulfur homeostasis (30). However, how the Cys-Tyr crosslink increases catalytic efficiency is not clear.

Three structures of substrate-bound CDO have also been reported: the first was a complex of human CDO with Cys (31), since shown to be flawed (32), the second a complex of rat CDO with a putative cysteine persulfenate formed in the active site (33), and the third a 1.45 Å resolution complex of rat CDO with Cys bound (PDB entry 4iew) (32). Our recent study showed that at pH $\leq$ 5.0 Cys does not bind to the active site, from pH 5.5 to 7.0 Cys-persulfenate is formed in the active site and at pH values of 8.0 and 9.0 Cys binds to the active site but oxygen does not bind and no persulfenate forms (32). We suggested that Cys does not bind below pH 5.5 due to the protonation of the Cys thiol or amino group, and that the pKa for the catalytic Tyr157 is between 7- 8. Tyr157 is part of a Ser153-His155-Tyr157 catalytic triad, and is centrally located close to both oxygen atoms of the persulfenate and is thought to serve as a catalytic acid/base (32). Whether it serves as a catalytic acid or base is not known, as there are different reports on the optimal active pH (recently summarized (32)). A recent study showed that persulfenate is indeed formed in the active site of crystalline enzyme (34), but that at pH 6.1 the crystalline enzyme does not release Cys-persulfenate/CSA from the crystal into the surrounding AML (34). That the crystalline enzyme does not complete the catalytic cycle is fortuitous in that the persulfenate is captured in the active site, but it could also be taken to mean that the persulfenate is off-pathway. However, that the Cys-persulfenate is formed enzymatically in the active site shows that it is energetically accessible (34), despite

quantum mechanical calculations suggesting it is not (23,35-37). One caveat of these calculations is the use of the flawed human CDO Cys-bound complex (32) and a different result might be obtained when using as a starting model the high-resolution Cys-bound complex of rat CDO (PDB entry 4iew) (32). Regardless, the question remains if the Cys-persulfenate complex represents an off-pathway inhibition-complex or an on-pathway intermediate. As the persulfenate only forms at pH ≤7, understanding the true pH profile of Cysteine dioxygenase could help elucidate this. Since the 2013 study on the pH dependence of persulfenate formation (32), three studies by two independent groups have been published which show decreased CSA production at pH <7.0 (3,34,38). This means that stabilized persulfenate formation occurs pH values at which catalysis does not occur, consistent with the persulfenate being an off-pathway dead-end complex that only occurs at low pH values (27), but not ruling out that it could still be an intermediate in the reaction at higher pH-values where the enzyme is active. In other words, it could be stuck at low pH values because a deprotonated Tyr157 is required to continue the catalytic cycle.

Here we investigate the structural role of the catalytic Tyr157 and the Cys93-Tyr157 crosslink by solving 17 structures of the C93A and Y157F variants in the presence and absence of Cys, as well as wild type CDO and each of the variants in the presence of the competitive inhibitors homocysteine and azide.

Results

In all, we describe in this report 17 new CDO structures at resolutions between 1.25 and 1.65 Å. Of these, 14 are of C93A and Y157F CDO either alone or in the presence of the Cys substrate or the inhibitor homocysteine; the other three structures are of wild-type CDO in the presence of homocysteine or azide. The two target pH values of 6.2 and 8.0 were chosen to assess possible pH-dependent differences in active site chemistry, as based on our previous study Cys-soaked wild type CDO crystals formed Cys-persulfenate at pH values from 5.5 to 7.0, but at pH=8 yielded a complex with only Cys present and no evidence of oxygen binding or reaction (39).

For clarity, in this presentation we have divided the remainder of the Results into sections treating groups of structures providing similar information.

The C93A and Y157F unliganded active sites

As purifying pure non-crosslinked mammalian CDO has been difficult, to gain insight into the structure of the non-crosslinked form of mammalian CDO we are using the C93A and Y157F mutants. The Y157F mutant can provide insight into how the unreacted Cys93 might pack with the phenyl group of Tyr157, and the C93A mutant can provides the best available control for whether active site changes seen in the Y157F form are due to the loss of the crosslink as opposed to being due to the loss of the Y157 hydroxyl group itself. A C93G mutation that would mimic the naturally occurring CDO enzymes that are uncrosslinked would not be as informative, because it would completely remove the non-covalent collision that should be exist between C93 and Y157 in the uncrosslinked form.

In the absence of a substrate/inhibitor soak, at either pH 8.0 or pH 6.2 the active sites of both C93A and Y157F contain very strong spherical $2F_o$-F_c map peaks (~17 ρ_{rms}; Figure 4.1A-C) ~2.2 Å away from the iron, that are of lower electron density than the iron (~ 39 ρ_{rms}), but slightly higher electron density than the sulfurs of Met and Cys residues (< 16 ρ_{rms}). Having more electrons than sulfur but fewer than iron, a chloride ion is consistent with the density and the chemistry and refines to reasonable temperature factors. Chloride is also known to bind to iron (24,25,40), and is present at ~15 mM in the crystallization conditions. The iron is five coordinate, with an additional water/hydroxide molecule bound to the iron in both Y157F and C93A. Interestingly, the additional water/hydroxide molecule present binds to a different site for the Y157F and C93A variants of CDO (Figure 4.1A-C and Figure 4.2).

For unsoaked C93A at pH 6.2 and 8.0, the Tyr157-OH shifts ~ 1.0 Å away from the iron and residue 93, relative to the cross-linked wild-type enzyme. The iron-bound water and chloride are fully occupied, and both interacting with with Tyr157, with the chloride receiving a ~3.5 Å hydrogen bond from Tyr157-OH (Figure 4.1A-

B). The shift of Tyr157 is not accompanied by a shift in the Ser153 and His155 residues of the catalytic triad and this causes His155 to no longer be well positioned to form a hydrogen bond with Tyr157 (Figure 4.2).

For unsoaked Y157F, clear density exists for a fully occupied chloride and a ~30% occupied water that also coordinates with the iron, and interacts with C93 (Figure 4.1C). The chloride is slightly shifted from the position in which it binds in C93A CDO, and the additional coordinating water is at a completely different position (Figure 4.2). Associated with the missing hydroxyl of residue 157, the Arg60 side chain also shifts to adopt an different conformation that we will refer to as position II, in which one of its guanidino NH atoms hydrogen bonds with Tyr58 and a water that in turn hydrogen bonds (at a typical 3.3 Å distance(41)) to the chloride (Figure 4.1C). This conformation of Arg60 has been seen previously in CDO forms having a disulfide modification of Cys164 (20,31) and is not compatible with Cys binding (32). These changes in Y157F CDO, represent a substantial variation of the active site hydration pattern compared with wild type cross-linked CDO or C93A (Figure 4.2),

To assess whether the chloride binding seen in the mutated CDOs was due to the iron being present in the ferric state, we soaked C93A and Y157F crystals with dithionite and methyl viologen. Although many crystals did not survive this treatment, crystals of Y157F soaked with dithionite at both pH 8 and 6.2 diffracted to 1.25 Å and allowed for a definitive structure determination. At this high resolution, the density was interpretable and showed that at both pH values, the active site largely contained chloride-bound in the same way as in the unsoaked crystals. The only difference in the active site, was the presence of some extra density which we were able to interpret as being due to thiosulfate ($S\text{-}SO_3$), a decomposition product of dithionite (42), bound at ~30% occupancy at pH 8.0 (Figure 4.1D), and at ~15% occupancy at pH 6.2 (Figure 4.1E). The high resolution was essential for being able to interpret the structure, as at lower contour levels it was possible to see discrete electron density for the oxygen atoms of the low occupancy thiosulfates. The thiosulfates bind with the terminal sulfur

ligating the iron at roughly the same position as the chloride, and with the SO_3 moiety hydrogen bonding with Arg60 sitting in position II.

Azide binding to CDO

Azide, which causes 50% inhibition of CDO at 1.4 mM (15), was soaked into crystals of wild-type, C93A and Y157F variants of CDO at 100 mM. Based on data sets at 1.5 Å resolution or better, we saw clear azide binding in the active site of wild-type CDO at pH 6.2 (Figure 4.1F), but no evidence for binding in the C93A or Y157F variants. The structure of azide-bound wild type CDO reveals that azide, unlike NO or CN, (2;3) does not act as a molecular oxygen mimic, but rather binds by ligating the iron at the same site as the same site seen above for chloride binding to the mutants; it also triggers the binding of an additional water/hydroxide to the iron (Figure 4.1F). The azide also interacts with the Tyr157-OH though the iron-proximal nitrogen. When wild type CDO crystals were soaked with both 100 mM azide and 100 mM Cys, the Cys out competed azide to give a result identical to that obtained based on a soak with Cys alone.

C93A and Y157F in the presence of Cysteine

When crystals of C93A or Y157F are soaked with 100 mM Cys, at either pH=6.2 or 8.0, the Fe-bound chloride is displaced, and Cys binds to the active site without any indication of molecular oxygen binding or Cys-persulfenate formation. The Cys binding modes are different than was seen in wild type cross-linked CDO and the results for C93A and Y157F are described in the following paragraphs.

For C93A crystals soaked with Cys at pH 8.0, Cys clearly binds, and we estimate its occupancy as 90% based on a small residual peak at the chloride positions indicating that ~10% of the chloride remains bound (Figure 4.3A). The Cys-thiol and α-amino group chelate the iron in a bi-dentate fashion (2.35 and 2.3 Å, respectively). The overall binding mode is quite similar to that seen in wild-type CDO, except for a small shift of ~0.6 Å of the α-carboxylate that appears to be caused by the altered

position of the Tyr157 hydroxyl (Figure 4.4). The active site Arg60 and Tyr58 that interact with the α-carboxylate, shift a bit less (~0.3 Å) leading small variations in the hydrogen bonds between them and the α-carboxylate. Interestingly, with Tyr157 shifted away from the iron, in the Cys complex a new strong water site is present in between the iron and the Tyr157 hydroxyl, where it also hydrogen bonds to His155,. The temperature factors of this water are consistent with it being bound at the same level (~90 %) as Cys, The position of this water molecule roughly matches the position seen for the distal oxygen (i.e. the OE atom) of Cys-persulfenate (Figure 4.4).

For C93A crystals soaked with Cys at pH 6.2, Cys binds at ~65% occupancy, with chloride remaining at ~35% occupancy (Figure 4.3B). Cys again coordinates the iron via its thiol group (~2.4 Å) but the α-amino group is further from the iron (~2.6 Å) than was seen the pH 8.0 or wild-type structures. This shift in Cys is correlated with a slight shift in Tyr58 as well as Arg60 in its standard position. There is also evidence for a minor conformation (~35%) of the Arg60 side chain adopts a position not seen before that we will call position III. This new Arg60 position allows its NE atom to hydrogen bond to the Cys carboxylate as it is bound in this structure.

For Y157F at pH 8.0, Cys completely displaces the chloride but binds in two conformations with the major (70%) state coordinating the iron in the typical bi-dentate fashion (via its thiol and α-amino group) and the minor (~30%) one coordinating the iron with the thiol group only. The position adopted by the thiol in the minor conformation has not been observed previously for Cys binding, but is close to the site of chloride binding (Figure 4.3C and Figure 4.4). This minor conformation is plausibly caused by protonation of the α-amino group, which would disrupt its ability to coordinate the iron and explain why it is 3.2 Å away from the iron. Arg60 is fully occupied in its standard position, forming hydrogen bonds to each oxygen of the α-carboxylate of the major conformation and to just one in the minor conformation (Figure 4.3C). Tyr58 forms a shorter hydrogen bond with Cys in the minor conformation (2.5 Å) than it does in the major conformation (2.9 Å).

For Y157F crystals exposed to Cys at pH 6.2, Cys binds in two conformations with the minor conformation being similar to the minor conformation for the Y157F mutant, but with the major conformation being completely different than any binding mode seen previously (Figure 4.3D). Both the minor (30%) and major (70%) conformations bind to the iron via only the thiol, with the α-amino atom being 3.3 Å and 4.3 Å away from the iron, respectively (Figure 4.3D). The sulfur for both conformations sits in approximately the same position as the chloride in the un-soaked structures. Associated with two positions of Cys, Arg60 adopts a mix of position III (70%) and position I (30%). Tyr58 adopts a single conformation that is shifted ~0.7 Å relative to wild-type, and forms a short ~2.4 Å hydrogen bond to the Cys in the minor conformation and a ~3.0 Å hydrogen bond with Cys in the major conformation.

Consistent with what was seen above, when Y157F crystals are exposed to Cys and dithionite at pH 8.0, the result (Figure 4.3E) is virtually identical to that seen in the Cys-only soak (Figure 4.3C). No evidence for thiosulfate binding is seen, and the active site does not change under these reducing conditions, indicating that the iron was already present in the ferrous form.

C93A, Y157F and Wild-type CDO in the presence of Homocysteine

For C93A, Y157F and wild-type CDO, soaks with the competitive inhibitor homocysteine (Hcy) lead to very similar binding with the thiol binding the iron (~2.15 Å) and the α-amino group ~3.5 Å from the iron (Figure 4.5B-F). The thiol is bound roughly in the position of the chloride and so does not match the position normally occupied the the Cys thiol. Tyr58 shifts ~0.6 Å relative to its position in the Cys-bound active site, allowing it to hydrogen bond to the α-carboxylate of Hcy, and Arg60 shifts about ~ 1.0 Å relative to its position in the cysteine-bound form. Unlike the wild-type Cys-bound form (PDB entry 4IEV) (39), only the NE atom forms a hydrogen bond (~3.2 Å) with the carboxylate of Hcy. The Arg60-NH1 atom does not form an H-bond with substrate, and but interacts with a second water molecule. Tyr157 is interacts with the Sγ and α-amino of Hcy (Figure 4.5B-F).

Of all of the Y157F, C93A and wild-type structures at pH 6.2 and pH 8.0 in the presence of Hcy, only the wild-type structure at pH 8 shows a unique conformation at 50% occupancy (Figure 4.5A), which we surmise is related to the deprotonation of the α-amino group of Hcy. In this complex, the Hcy α-amino group forms an even shorter H-bond to Tyr157-OH than did the Cys α-amino group (~2.8 vs ~3.0 Å).

Discussion

The positioning of Tyr157 is critical for iron coordination

Disruption of the Cys-Tyr crosslink by mutation of either Cys93 to Ala or Tyr157 to Phe results in phenyl ring of residue 157 undergoing a nearly identical shift of ~ 1.0 Å relative to cross-linked enzyme (Figure 4.4). The main cause of the shift is to prevent a clash between the Cβ of residue 93 and the carbon at the ortho position of residue 157. That the shift is nearly identical in both mutants allows us to conclude that the Sγ atom of Cys93 does not push away residue 157 any more than does the Ala, and that the shift seen is representative of what would exist in non-crosslinked wild-type CDO which combines a Cys at position 93 with a Tyr at position 157.

In wild type cross-linked CDO without substrate bound, the Tyr157 hydroxyl makes a rather short (~2.6 Å) hydrogen bond to the iron bound water, and the importance of this interaction for maintaining that coordination is seen in that for both mutants the iron coordination present in the crystal shifts from the three-His, one-water coordination, to a pentacoordinate three-His, one-chloride, one-water coordination where the chloride binds in roughly the same place as the water. Because O···Cl distances are on the order of 3.0 to 3.7 Å (41), the native position of the Tyr157 hydroxyl (2.6 Å from the Fe-bound water) presumably excludes the chloride from binding to the wild-type crosslinked enzyme. The Fe-Cl bond in SyrB2 is 2.44 Å, longer than the average Fe (II)-Cl bond length of 2.3 Å for a six-coordinate iron, (43) and the ~2.2 Å Fe-Cl bond in the non-crosslinked CDO variants.

Tyr157 is important for proper Cys binding and persulfenate formation

Exposing crystals of C93A and Y157F variants of CDO to Cys results in direct coordination of Cys to the active-site iron, consistent with changes in EPR signal observed with the addition of Cys to the C93A and Y157F variants (3), or to non-crosslinked wild-type enzyme (28). In the crosslinked enzyme, Cys binds to the iron from pH 5.5 to 9.0 in the same mode via its α-amino and thiol groups, indicating that when bound in the active site the α-amino group is deprotonated over this entire range (32). Since a typical α-amino group pKa is around 8 or 9, the when bound to the active site of CDO this pKa is substantially lowered to be near 4.5. The structures presented here define Tyr157 as a key factor that modulates this pKa. Based upon the relative occupancies of Cys molecules bound in the active site, when the Tyr157-OH is removed (i.e. Y157F) the α-amino group of Cys is 30% protonated at pH 8 and completely protonated at pH 6.2, indicating a pKa similar to that of a free amino acid. When the Tyr157-OH is shifted (i.e. in the C93A mutant) the weaker binding of Cys at pH 6.2 and the longer Fe-$N_{(Cys)}$ distance implies it is partly protonated that that pH, meaning that even a shifted Tyr157-OH is somewhat able to modulate the α-amino group pKa. That the Cys93-Tyr157 crosslink is especially important for facilitating the Fe-$N_{(Cys)}$ coordination is consistent with EPR results showing that the most pronounced increase in g strain upon the disruption of the C93-Y157 crosslink by mutation is along the Fe-N(Cys) bond (g1 axis) (3).

Furthermore, the ~1.0 Å shift in Y157 upon disruption of the crosslink causes a misalignment of the catalytic triad, as His155 does not change conformation and so the His155 to Tyr157 hydrogen bonding interaction is elongated from ~2.9 to ~3.1 Å and its linearity is decreased by about ~20˚ (Figure 4.2). The substantial nature of this disruption is underscored by the binding of a new ordered water that accepts a hydrogen bonds from His155 in the Cys complex of C93A. This ordered water blocks what would be the dioxygen binding site in the wild type crosslinked enzyme (2,3,5,32,33). This competition with dioxygen may contribute to the lower catalytic activity of the uncrosslinked enzyme along with the direct impacts of the shifted

Tyr157 no longer being well placed to directly interact with the molecular oxygen, or to be activated by the catalytic triad. A recent steady-state kinetics data showed that a H155A variant of CDO has no detectable catalytic activity (3), consistent with the importance of the His155-Tyr157 hydrogen bond.

Inhibition by homocysteine and azide

In Alzheimer's disease high levels of Hcy are observed together with elevated levels of cysteine (14), raising the possibility that CDO inhibition by Hcy, known to have an IC50 of 6.5 mM (16), plays a role. Although previous attempts to crystallize human CDO in the presence of Hcy were unsuccessful (31), we were able through soaking experiments to capture Hcy in the active site of wild-type, C93A and Y157F CDO. Hcy is not a substrate for CDO, and it has been suggested that there must be an energetic reason why the iron(II)CDO-Hcy complex either does not bind oxygen or does not react with it (4). The solved complexes make it clear that the inability of CDO to turnover Hcy is simply due to its thiol binding in an incorrect location that also blocks the oxygen binding site (as inferred from the position of the persulfenate oxygens (32,33)). Noteworthy is that in none of these complexes, does the Hcy α-amino group also coordinate the iron (Figure 4.5). The reason for the difference in binding appears to be related to the greater separation between the Hcy α-amino group and its thiol group, so that they cannot simultaneously bind to the iron in the same way that they do for Cys. The very similar binding of Hcy to the non-crosslinked CDO forms may be of physiological relevance in that the inhibition of non-crosslinked CDO by high levels of Hcy could, in addition to directly preventing Cys conversion to CSA production, also inhibit CDO crosslink formation.

In contrast to this, the inhibition by azide, based on the crystal structures done here, only appears to occur for the crosslinked enzyme, since no complexes could be formed in the C93A or Y157F crystals. It is interesting to note that the azide inhibition studies, defining the Ki of 1.4 mM (15), were carried out before the discovery of the Cys-Tyr crosslink (21), and were done on an as-isolated population of CDO that

presumably contained a mix of both forms. This means that Ki value may be underestimated. Given recent methods for obtaining pure populations of the non-crosslinked (28) and crosslinked (27,29) wild-type enzyme, it will be of interest to directly measure the Ki values for various inhibitors for the crosslinked and non-crosslinked forms.

Materials and Methods

Expression, purification and crystallization

Expression, purification and crystallization of wild-type, C93A and Y157F recombinant rat CDO was done as described previously for wild-type rat CDO (32,33,44,45), with seeding being especially important for consistent (~90% of drops) nucleation and crystal growth for the Y157F variant, which without seeding crystallizes in only ~5% of drops. The enzyme was stored at 6-10 mg/mL in 10 mM Tris pH=7.4 buffer and crystallized at room temperature in hanging drops formed from 0.5 μL seed stock in reservoir buffer, 1 μL protein stock and 1 μL of reservoir buffer composed of 0.1 M tri-sodium citrate pH=5.6, 24% PEG 4000, and 0.15 M ammonium acetate (final pH=6.2) (32). The crystals of all variants were isomorphous with those previously described (32,33,44), having space group $P4_32_12$ and one chain in the asymmetric unit.

Crystal soaks and data collection

Soaks were done at room temperature using the reservoir condition as an artificial mother liquor (AML). For data collection, crystals were flash frozen by plunging into liquid nitrogen. For cysteine, homocysteine and azide soaks, crystals were incubated for 10 min in solutions with 100 mM cysteine or homocysteine added to AML that had its pH adjusted to a final value of 6.2 or 8.0. Ligand-free structures at pH 6.2 were scooped directly from the hanging drop. As previously (32), for ligand-free structures at pH 8.0, crystals were incubated for 10 min in solutions that contained 100 mM alaninine to serve as a buffer that was similar to cysteine. For preparing the dithionite-reduced form, crystals were incubated for 45 min under argon in a solution

of 5 mM dithionite and 0.5 mM methylviologen in AML with or without 100 mM Cys which had been degassed and flushed with argon. The blue color of reduced methyl viologen was present during the course of the incubation, indicating that the reducing power of the solution was not exhausted.

Synchrotron data sets were collected at beamlines 5.0.1, 5.0.2, and 5.0.3 at the Advanced Light Source (Lawrence Berkeley National Laboratory). Data were processed and scaled using iMosflm (46) and Aimless (47). For all refinements, the same 10% of the data as used previously (32,33) were flagged for use in Rfree. Data collection statistics are in Tables 4.1, 4.2 and 4.3.

Crystallographic refinement

For each of the 17 structures reported here, 8 different initial refinement strategies were tested using either the unique unit cell constants for individual crystals (within the ranges a=b=57.6-58.2 Å and c=121.7-122.9 Å) or using a=b=57.60 Å and c=122.40 Å. Six of the refinement strategies tested were similar to the automated protocol used previously (32,48), and began with the 1.5 Å resolution unliganded CDO model (PDB code 2b5h) or with the 1.55 Å resolution unliganded CDO model (PDB code 4ieo) either with isotropic temperature factors, with one TLS group defined by the entire chain, or with individual anisotropic temperature factors. Two additional refinement strategies began with PDB code 4ieo with waters retained and a simple minimization either with one TLS group defined by the entire chain, or with individual anisotropic temperature factors. The anisotropic B-factors were only accepted if their use resulted in an Rfree improvement of at least 1% compared with the TLS refinements. The best refinement strategy used the same unit cell for all refinements (a=b=57.60 Å and c=122.40 Å) with 4ieo as the starting model with the automated protocol. For refinement against all but two of the datasets, the best strategy used TLS. The anisotropic B-factors were used for Y157F dithionite-treated structures at pH 6.2 and 8.0.

After the automated stage, each refinement was completed using manual rebuilding in Coot (49) and PHENIX-refine minimization. Some water sites were removed and others were added using standard criteria (>1 ρ_{rms} intensity in the 2Fo −Fc map, >2.4 Å distance from nearest contact), and all difference map peaks above 5*ρ_{rms} were checked. For those refinements using TLS, the temperature factors of the iron were refined anisotropically, and this resulted in an improvement of ~0.5% in R_{free}. Molprobity (50) was used to find problems with model geometry. Refinement statistics are in Tables 4.1-3.

Acknowledgements

We thank Dale Tronrud, Rick Cooley and John Dominy for useful discussions and/or help with methods. This project was supported in part by Grant NIH-DK-056649 to MHS and PAK from the National Institute of Diabetes and Digestive and Kidney Diseases. Synchrotron data were collected at the Advanced Light Source supported by Contract DE-AC02-98CH10886 from the Office of Basic Energy Sciences of the U.S. Department of Energy.

Abbreviations

H-bonds: hydrogen bonds, CDO: Cysteine Dioxygenase, AML: artificial mother liquor, Cys: Cysteine, CSA: cysteine sulfinic acid.

Table 4.1. Data collection and Refinement Stats for nosoak, dt and azide CDOs[a]

	C93A no soak pH 8.0	**C93A no soak pH 6.2**	**Y157F no soak pH 6.2**	**Y157F dithionite pH 8.0**	**Y157F dithionite pH 6.2**
Data collection					
Resolution (Å)	41-1.35 (1.37-1.35)	39-1.35 (1.42-1.35)	39-1.35 (1.37-1.35)	26-1.25 (1.27-1.25)	29-1.25 (1.27-1.25)
Unique Obs.	45533 (1757)	40379 (5123)	40658 (995)	53790 (1261)	49712 (1202)
Multiplicity	23.3 (9.1)	23.4 (10.2)	27.5 (22.5)	22.2 (6.8)	24.0 (7.6)
Completeness	97.9 (78.6)	86.2 (76.8)	88.6 (45.1)	92.3 (45.2)	85.4 (41.9)
$\langle I/\sigma \rangle$	22.5 (1.0)	12.7 (0.6)	30.1 (2.8)	27.1 (1.4)	29.3 (1.4)
R_{meas}[b] (%)	10.3 (257)	16.8 (602)	8.9(129)	7.0 (129)	7.7 (155)
R_{pim} (%)	2.1 (82.4)	3.4 (181)	1.7 (26.4)	1.4 (47.9)	1.5 (55.1)
$CC_{1/2}$[c] (%)	1.00 (0.29)	1.0 (0.11)	1.0 (0.77)	1.0 (0.56)	1.0 (0.47)
Refinement					
R_{cryst} / R_{free} (%)	16.1/19.1	16.7/20.2	14.9/17.8	12.1/16.2	12.1/16.4
No. Obs	45445	39568	40635	53704	49645
No. residues	186	186	186	186	186
No. waters	333	143	311	354	344
No. atoms	1952	1717	1873	1935	1938
rmsd angles (°)	1.30	1.27	1.26	1.25	1.18
rmsd lengths (Å)	0.014	0.011	0.013	0.011	0.011
φ,ψ-favored (%)[d]	99.5	98.97	98.96	98.97	98.98
φ,ψ-outliers (%)[d]	0	0	0	0	0
$\langle B \rangle$ protein (Å^2)	17.9	23.3	16.6	16.7	16.3
$\langle B \rangle$ ligand (Å^2)	18.3	24	11.4	14.9	16.5
PDB code	4PIX	4PIY			

[a] All refinements used space group $P4_32_12$ with a=b=57.60 Å and c= 122.40 Å. Numbers in parentheses refer to the highest resolution bin.
[b] R_{meas} is the multiplicity-weighted merging R-factor (51)
[c] $CC_{1/2}$ is the correlation between two datasets each based on half of the data as defined in Karplus & Diederichs.(48)
[d]Ramachandran statistics as defined by Molprobity (50)

Table 4.2. Data collection and Refinement Statistics for Cys-complexed C93A and Y157F CDOs[a]

	C93A Cys pH 8.0	**C93A Cys pH 6.2**	**Y157F Cys pH 8.0**	**Y157F Cys pH 6.2**	**Y157F Cys + dt pH 8.0**
Data collection					
Resolution (Å)	34-1.40 (1.42-1.40)	34-1.40 (1.42-1.40)	34-1.55 (1.58-1.55)	28-1.35 (1.37-1.35)	34-1.55 (1.58-1.55)
Unique Obs.	37823 (1109)	41405 (2011)	31437 (1539)	45246 (1785)	31232 (1519)
Multiplicity	28.1 (26.4)	23.7 (13.4)	25.9 (16.3)	23.0 (8.6)	26.0 (16.3)
Completeness	91.8 (54.1)	100 (100)	100 (100)	97.8 (80.7)	100 (100)
$\langle I/\sigma \rangle$	22.5 (1.8)	11.3 (0.8)	21.4 (0.6)	17.8 (0.7)	17.3 (0.6)
R_{meas} (%)	15.4 (366)	18.2 (275)	7.6 (596)	11.7 (276)	9.0 (500)
R_{pim} (%)	2.9 (71)	3.7 (74)	1.5 (147)	2.4 (92)	1.8 (123)
$CC_{1/2}$	1.00 (0.26)	1.0 (0.19)	1.0 (0.24)	1.0 (0.20)	1.0 (0.23)
Refinement					
R_{cryst} / R_{free} (%)	15.5/18.6	16.3/19.7	18.3/22.4	16.2/19.3	18.7/22.2
No. Obs	37752	41323	30548	45159	30399
No. residues	186	186	186	186	186
No. waters	302	299	143	319	133
No. atoms	1869	1870	1717	1907	1702
rmsd angles (°)	1.28	1.28	1.27	1.26	1.30
rmsd lengths (Å)	0.010	0.015	0.012	0.011	0.014
φ,ψ-favored (%)	98.96	98.96	98.45	98.98	98.96
φ,ψ-outliers (%)	0	0	0	0	0
$\langle B \rangle$ protein (Å^2)	19.8	20.1	46.2	19.8	46.5
$\langle B \rangle$ ligand (Å^2)	13.6	15.2	32.9	21.7	37.6
PDB code					

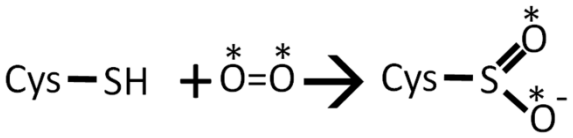
[a] See Table 4.1 footnotes for definitions related to Table entries

Table 4.3. Data collection and Refinement Statistics for Hcy-complexed CDOs[a]

	Wild-type Hcy pH 8.0	**Wild-type Hcy pH 6.2**	**C93A Hcy pH 8.0**	**C93A Hcy pH 6.2**	**Y157F Hcy pH 8.0**	**Y157F Hcy pH 6.2**	**Wild-type azide pH 6.2**
Data collection							
Resolution (Å)	41-1.35 (1.37-1.35)	34-1.40 (1.48-1.40)	34-1.30 (1.32-1.30)	42-1.60 (1.69-1.60)	33-1.65 (1.68-1.65)	33-1.40 (1.42-1.40)	34-1.50 (1.58-1.50)
Unique Obs.	43919 (1521)	42028 (6012)	43282 (896)	28457 (4040)	25597 (1236)	39877 (1434)	33827 (4812)
Multiplicity	23.8 (10.3)	7.6 (5.6)	26.7 (17.9)	13.8 (13.7)	27.4 (22.8)	23.3 (9.0)	11.8 (11.9)
Completeness	94.9 (68)	100 (100)	84.2 (36.2)	99.6 (99.0)	98.9 (95.2)	94.7 (67.8)	100 (100)
$\langle I/\sigma \rangle$	23.7 (1.0)	12.5 (0.6)	21.3 (1.2)	12.1 (1.1)	26.3 (2.0)	24.7 (1.1)	8.6 (0.6)
R_{meas} (%)	8.0 (245)	7.6 (287)	10.5 (251)	27.9 (495)	9.1 (199)	8.3 (253)	18.2 (538)
R_{pim} (%)	1.6 (75)	2.7 (119)	2.0 (57)	7.4 (132)	1.7 (40)	1.7 (82)	5.3 (155)
$CC_{1/2}$	1.00 (0.26)	1.0 (0.18)	1.00 (0.33)	1.0 (0.26)	1.00 (0.60)	1.0 (0.36)	1.0 (0.17)
Refinement							
R_{cryst} / R_{free} (%)	16.2/19.1	17.1/19.8	15.7/18.6	16.4/20.9	16.9/21.0	16.2/19.2	17.2/21.0
No. Obs	43838	41879	43245	27819	25071	39392	33667
No. residues	186	186	186	186	186	186	186
No. waters	280	247	302	279	182	281	214
No. atoms	1914	1823	1907	1856	1749	1884	1785
rmsd angles (°)	1.36	1.3	1.27	1.28	1.33	1.32	1.31
rmsd lengths (Å)	0.011	0.011	0.010	0.012	0.013	0.014	0.012
φ,ψ-favored (%)	99	99	99	98.5	99.5	99	99.5
φ,ψ-outliers (%)	0	0	0	0	0	0	0
$\langle B \rangle$ protein (Å^2)	23.3	27.5	20.3	24.3	39.9	21.4	27.7
$\langle B \rangle$ ligand (Å^2)	33.1	42.1	33.2	37.8	29.7	22.3	47.4
PDB code		4PIZ		4PIY			4PJY

[a] See Table 4.1 footnotes for definitions related to Table entries

Scheme 4.1. Reaction Catalyzed by Cysteine Dioxygenase



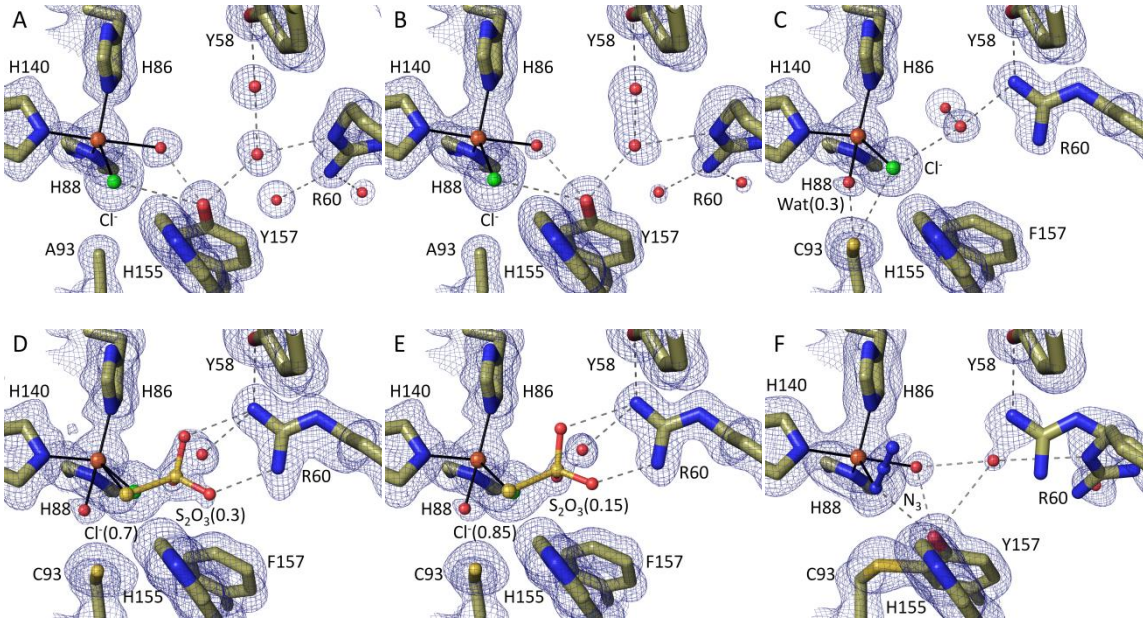


Figure 4.1. Active site density of non cross-linked variants shows a Fe-bound chloride. (A) C93A in AML at pH 8.0, (B) C93A in AML at pH 6.2, (C) Y157F in AML at pH 6.2, (D) Y157F with dithionite at pH 6.2, (E) Y157F with dithionite at pH 8.0, and (F) wild-type with azide-bound at pH 6.2, with all $2F_o$-F_c density contoured at 1.4 ρ_{rms}.

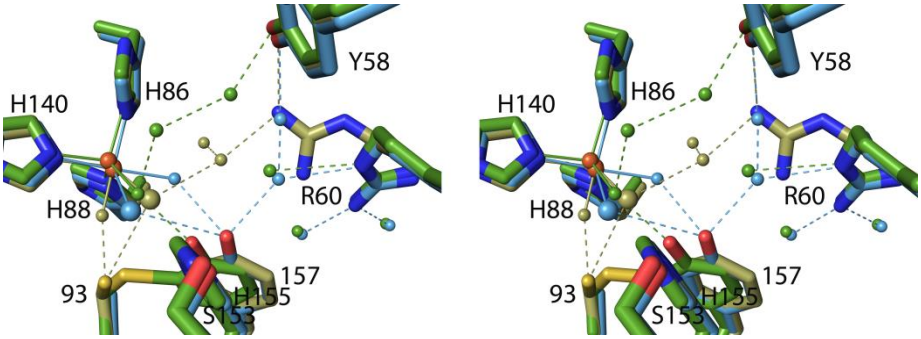


Figure 4.2. Stereo view of unsoaked crosslinked and non crosslinked active sites. Ligand-free crosslinked CDO (PDB entry 2b5h)(20) is shown with green carbons/waters; Chloride-bound C93A is shown (blue carbons/water/chloride with the iron-bound chloride represented as a larger sphere than the water molecules; Chloride-bound Y157F (gold carbons/water/chloride with the chloride sphere larger). The Tyr157 shift upon disruption of the crosslink is clear, with the Ser153-His155 shown no longer well positioned to hydrogen bond to Tyr157.

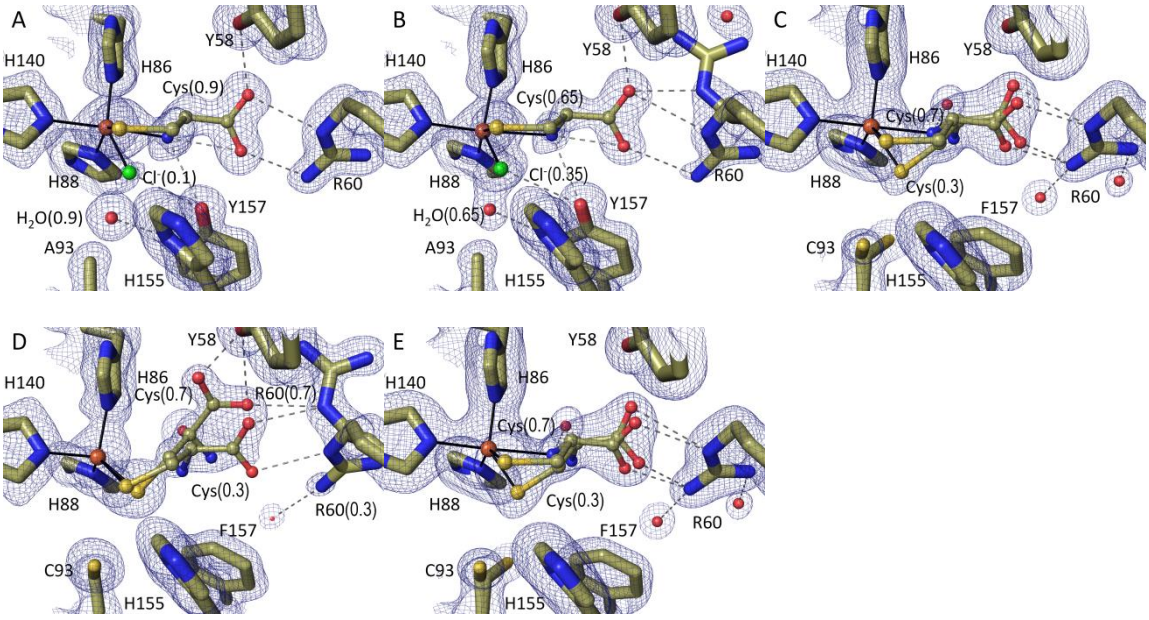


Figure 4.3. Cysteine-complexes of non cross-linked variants. (A) C93A with Cys at pH 8.0, (B) C93A with Cys at pH 6.2, (C) Y157F with Cys at pH 8.0, (D) Y157F with Cys at pH 6.2, and (E) Y157F with Cys and dithionite at pH 8.0 with all $2F_o$-F_c density contoured at 1.4 ρ_{rms}.

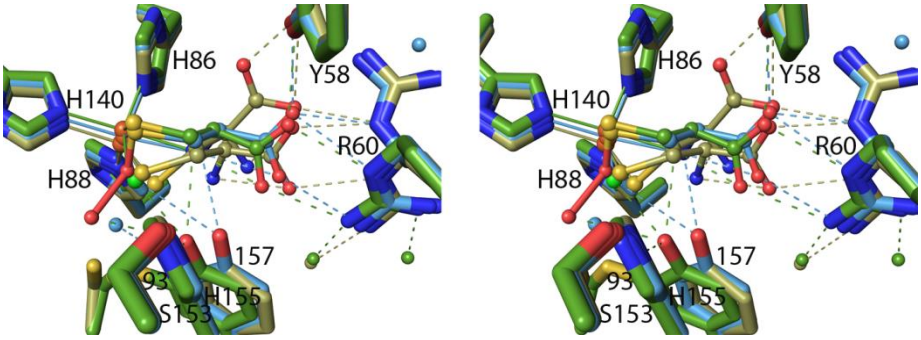


Figure 4.4. Stereo view of the four binding modes of Cys. The four different binding modes of Cys are captured in the active site with wild-type Cys-persulfenate (green carbons/waters) shown with the thiol and α-amino groups coordinating the iron and the α-carboxyl group interacting with Tyr58, Arg60, and Tyr157. A similar binding mode is seen in Cys-bound C93A (blue carbons/waters) where Cys is still coordinating the iron in a bi-dentate fashion with the α-amino and thiol groups, but with an additional water H-bonding to His155 near the position of the OD atom of the persulfenate, with the α-carboxyl group of Cys shifted along with Tyr157, and Arg60 adopting position III. The Cys-bound Y157F (gold carbons/waters), with one alternative conformations common to C93A at pH 6.2, and a unique conformation with both oxygen atoms of the α-carboxyl group interacting with the hydroxyl of Tyr58. In both of these alternative conformation, Cys is binding to the iron via the thiol but not the α-amino group.

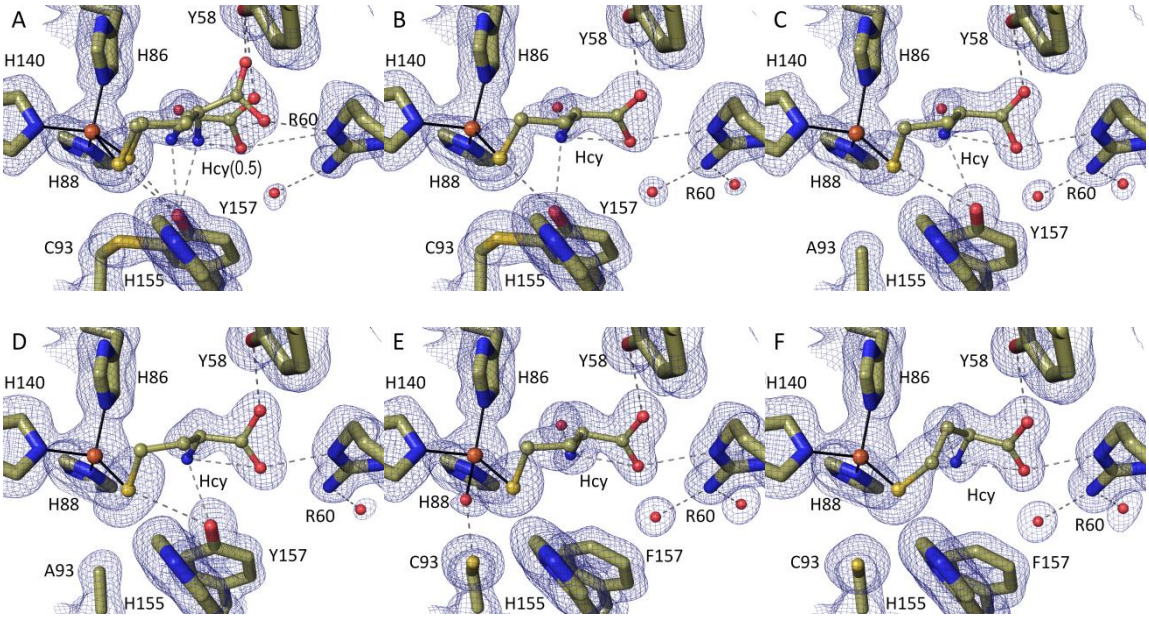


Figure 4.5. Homocysteine complexes of wild-type, C93A and Y157F Cysteine dioxygenase. (A) Wild-type CDO with homocysteine at pH 8.0, (B) Wild-type CDO with homocysteine at pH 6.2, (C) C93A with homocysteine at pH 8.0, (D) C93A with homocysteine at pH 6.2, (E) Y157F with Homocysteine at pH 8.0, and (F) Y157F with homocysteine at pH 6.2, with all $2F_o$-F_c density contoured at 1.4 ρ_{rms}.

References

1. Stipanuk, M. H., Simmons, C. R., Karplus, P. A., and Dominy, J. E., Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. Amino Acids 41, 91-102
2. Pierce, B. S., Gardner, J. D., Bailey, L. J., Brunold, T. C., and Fox, B. G. (2007) Characterization of the nitrosyl adduct of substrate-bound mouse cysteine dioxygenase by electron paramagnetic resonance: electronic structure of the active site and mechanistic implications. Biochemistry 46, 8569-8578
3. Li, W., Blaesi, E. J., Pecore, M. D., Crowell, J. K., and Pierce, B. S. (2013) Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O(2) coupling efficiency in mouse cysteine dioxygenase. Biochemistry 52, 9104-9119
4. Tchesnokov, E. P., Wilbanks, S. M., and Jameson, G. N. (2012) A strongly bound high-spin iron(II) coordinates cysteine and homocysteine in cysteine dioxygenase. Biochemistry 51, 257-264
5. Gardner, J. D., Pierce, B. S., Fox, B. G., and Brunold, T. C. (2010) Spectroscopic and computational characterization of substrate-bound mouse cysteine dioxygenase: nature of the ferrous and ferric cysteine adducts and mechanistic implications. Biochemistry 49, 6033-6041
6. Stipanuk, M. H., Hirschberger, L. L., Londono, M. P., Cresenzi, C. L., and Yu, A. F. (2004) The ubiquitin-proteasome system is responsible for cysteine-responsive regulation of cysteine dioxygenase concentration in liver. Am J Physiol-Endoc M 286, E439-E448
7. Dominy, J. E., Jr., Hirschberger, L. L., Coloso, R. M., and Stipanuk, M. H. (2006) Regulation of cysteine dioxygenase degradation is mediated by intracellular cysteine levels and the ubiquitin-26 S proteasome system in the living rat. Biochem J 394, 267-273

8. Park, S., and Imlay, J. A. (2003) High levels of intracellular cysteine promote oxidative DNA damage by driving the Fenton reaction. Journal of Bacteriology 185, 1942-1950
9. Olney, J. W., Zorumski, C., Price, M. T., and Labruyere, J. (1990) L-Cysteine, a Bicarbonate-Sensitive Endogenous Excitotoxin. Science 248, 596-599
10. Hogins, J., Crawford, D. C., Zorumski, C. F., and Mennerick, S. (2011) Excitotoxicity Triggered by Neurobasal Culture Medium. PLoS One 6
11. Stipanuk, M. H. (2004) Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annu Rev Nutr 24, 539-577
12. Heafield, M. T., Fearn, S., Steventon, G. B., Waring, R. H., Williams, A. C., and Sturman, S. G. (1990) Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. Neurosci Lett 110, 216-220
13. Bradley, H., Gough, A., Sokhi, R. S., Hassell, A., Waring, R., and Emery, P. (1994) Sulfate metabolism is abnormal in patients with rheumatoid arthritis. Confirmation by in vivo biochemical findings. J Rheumatol 21, 1192-1196
14. McCaddon, A., Hudson, P., Hill, D., Barber, J., Lloyd, A., Davies, G., and Regland, B. (2003) Alzheimer's disease and total plasma aminothiols. Biol Psychiat 53, 254-260
15. Chai, S. C., Bruyere, J. R., and Maroney, M. J. (2006) Probes of the catalytic site of cysteine dioxygenase. J Biol Chem 281, 15774-15779
16. Chai, S. C., Jerkins, A. A., Banik, J. J., Shalev, I., Pinkham, J. L., Uden, P. C., and Maroney, M. J. (2005) Heterologous expression, purification, and characterization of recombinant rat cysteine dioxygenase. J Biol Chem 280, 9865-9869
17. Ueki, I., Roman, H. B., Valli, A., Fieselmann, K., Lam, J., Peters, R., Hirschberger, L. L., and Stipanuk, M. H. (2011) Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize

taurine and an increased catabolism of cysteine to hydrogen sulfide. Am J Physiol Endocrinol Metab 301, E668-684

18. Brait, M., Ling, S., Nagpal, J. K., Chang, X., Park, H. L., Lee, J., Okamura, J., Yamashita, K., Sidransky, D., and Kim, M. S. (2012) Cysteine dioxygenase 1 is a tumor suppressor gene silenced by promoter methylation in multiple human cancers. PLoS One 7, e44951
19. Shaker, M., Pascarelli, K. M., Plantinga, M. J., Love, M. A., Lazar, A. J., Ingram, D. R., von Mehren, M., Lev, D., Kipling, D., and Broccoli, D. (2014) Differential expression of cysteine dioxygenase 1 in complex karyotype liposarcomas. Biomark Cancer 6, 1-10
20. Simmons, C. R., Liu, Q., Huang, Q., Hao, Q., Begley, T. P., Karplus, P. A., and Stipanuk, M. H. (2006) Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. J Biol Chem 281, 18723-18733
21. McCoy, J. G., Bailey, L. J., Bitto, E., Bingman, C. A., Aceti, D. J., Fox, B. G., and Phillips, G. N., Jr. (2006) Structure and mechanism of mouse cysteine dioxygenase. Proc Natl Acad Sci U S A 103, 3084-3089
22. Crawford, J. A., Li, W., and Pierce, B. S. (2011) Single turnover of substrate-bound ferric cysteine dioxygenase with superoxide anion: enzymatic reactivation, product formation, and a transient intermediate. Biochemistry 50, 10241-10253
23. de Visser, S. P., and Straganz, G. D. (2009) Why do cysteine dioxygenase enzymes contain a 3-His ligand motif rather than a 2His/1Asp motif like most nonheme dioxygenases? J Phys Chem A 113, 1835-1846
24. Blasiak, L. C., and Drennan, C. L. (2009) Structural Perspective on Enzymatic Halogenation. Accounts Chem Res 42, 147-155
25. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S., and Walsh, C. T. (2006) Nature's inventory of halogenation catalysts: Oxidative strategies predominate. Chem Rev 106, 3364-3378

26. Kleffmann, T., Jongkees, S. A., Fairweather, G., Wilbanks, S. M., and Jameson, G. N. (2009) Mass-spectrometric characterization of two posttranslational modifications of cysteine dioxygenase. J Biol Inorg Chem 14, 913-921
27. Dominy, J. E., Jr., Hwang, J., Guo, S., Hirschberger, L. L., Zhang, S., and Stipanuk, M. H. (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. J Biol Chem 283, 12188-12201
28. Njeri, C. W., and Ellis, H. R. (2014) Shifting Redox States of the Iron Center Partitions CDO between Crosslink Formation or Cysteine Oxidation. Arch Biochem Biophys
29. Siakkou, E., Rutledge, M. T., Wilbanks, S. M., and Jameson, G. N. (2011) Capturing crosslink formation with enzymatic activity in cysteine dioxygenase. Biochim Biophys Acta
30. Dominy, J. E., Jr., Hwang, J., and Stipanuk, M. H. (2007) Overexpression of cysteine dioxygenase reduces intracellular cysteine and glutathione pools in HepG2/C3A cells. Am J Physiol Endocrinol Metab 293, E62-69
31. Ye, S., Wu, X., Wei, L., Tang, D., Sun, P., Bartlam, M., and Rao, Z. (2007) An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. J Biol Chem 282, 3391-3402
32. Driggers, C. M., Cooley, R. B., Sankaran, B., Hirschberger, L. L., Stipanuk, M. H., and Karplus, P. A. (2013) Cysteine dioxygenase structures from pH4 to 9: consistent cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH. J Mol Biol 425, 3121-3136
33. Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Jr., Begley, T. P., Stipanuk, M. H., and Karplus, P. A. (2008) A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. Biochemistry 47, 11390-11392

34. Souness, R. J., Kleffmann, T., Tchesnokov, E. P., Wilbanks, S. M., Jameson, G. B., and Jameson, G. N. (2013) Mechanistic implications of persulfenate and persulfide binding in the active site of cysteine dioxygenase. Biochemistry 52, 7606-7617
35. Aluri, S., and de Visser, S. P. (2007) The mechanism of cysteine oxygenation by cysteine dioxygenase enzymes. J Am Chem Soc 129, 14846-14847
36. Kumar, D., Thiel, W., and de Visser, S. P. (2011) Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. J Am Chem Soc 133, 3869-3882
37. Kumar, D., Sastry, G. N., Goldberg, D. P., and de Visser, S. P. (2012) Mechanism of S-oxygenation by a cysteine dioxygenase model complex. J Phys Chem A 116, 582-591
38. Fellner, M., Doughty, L. M., Jameson, G. N., and Wilbanks, S. M. (2014) A chromogenic assay of substrate depletion by thiol dioxygenases. Anal Biochem 459C, 56-60
39. Driggers, C. M., Ellis, H. R., and Karplus, P. A. (2012) Crystal Structure of Escherichia coli NADPH FMN reductase SsuE with and without bound FMN. Proceedings of the 17th International Symposium of Flavins and Flavoproteins. 17, 613-618
40. Witten, E. H., Reiff, W. M., Lazar, K., Sullivan, B. W., and Foxman, B. M. (1985) The Ferric Chloride-Alpha-Diamine System .3. X-Ray Crystallographic, Magnetic-Susceptibility, and Zero-Field and High-Field Mossbauer-Spectroscopy Investigation of [Fe(2,2'bpy)2cl2][Fecl4] - Slow Paramagnetic Relaxation and Magnetic-Ordering of Complex Bimetallic Salts. Inorganic Chemistry 24, 4585-4591
41. Steiner, T. (1998) Hydrogen-bond distances to halide ions in organic and organometallic crystal structures: Up-to-date database study. Acta Crystallogr B 54, 456-463

42. Kilroy, W. P. (1980) Anaerobic Decomposition of Sodium Dithionite in Alkaline-Solution. J Inorg Nucl Chem 42, 1071-1073
43. Goldsmith, C. R., Jonas, R. T., Cole, A. P., and Stack, T. D. (2002) A spectrochemical walk: single-site perturbation within a series of six-coordinate ferrous complexes. Inorg Chem 41, 4642-4652
44. Simmons, C. R., Hao, Q., and Stipanuk, M. H. (2005) Preparation, crystallization and X-ray diffraction analysis to 1.5 A resolution of rat cysteine dioxygenase, a mononuclear iron enzyme responsible for cysteine thiol oxidation. Acta Crystallogr Sect F Struct Biol Cryst Commun 61, 1013-1016
45. Simmons, C. R., Hirschberger, L. L., Machi, M. S., and Stipanuk, M. H. (2006) Expression, purification, and kinetic characterization of recombinant rat cysteine dioxygenase, a non-heme metalloenzyme necessary for regulation of cellular cysteine levels. Protein Expr Purif 47, 74-81
46. Leslie, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EAMCB newsletter on protein crystallography 26
47. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr 67, 282-292
48. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. Science 336, 1030-1033
49. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D 66, 486-501
50. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol 285, 1711-1733
51. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4, 269-275

Chapter 5

Crystal Structures of *Bacillus subtilis* Cysteine Dioxygenase and a *Ralstonia eutropha* Cysteine Dioxygenase Homolog

Camden M. Driggers, Steven J. Hartman, and P. Andrew Karplus

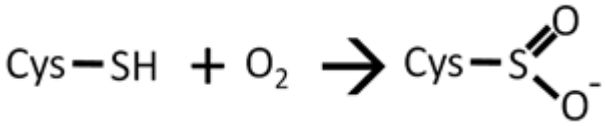
In preparation as a Structure Report for *Protein Science*

Abstract

In some bacteria, cysteine is converted to cysteine sulfinic acid by cysteine dioxygenase (CDO) using enzymes homologous to, but rather divergent (<25% sequence identity) from mammalian CDOs. Among bacterial proteins having this range of sequence similarity to mammalian CDO are some that conserve an active site Arg residue ("Arg-type" enzymes) and some having a Gln substituted for this Arg ("Gln-type" enzymes). To date no report on the structure of a bacterial CDO has been published. Here we describe a structure from each of these enzyme types by reporting on structures originally solved by structural genomics groups but not published: a *Bacillus subtilis* "Arg-type" enzyme that has cysteine dioxygenase activity (*Bs*CDO), and *Ralstonia eutropha* is "Gln-type" CDO homolog of uncharacterized activity (*Re*CDOhom). The *Bs*CDO active site is well conserved with mammalian CDO, and a cysteine complex captured in the active site confirms that the cysteine binding mode is also similar. The *Re*CDOhom structure reveals a new active site Arg residue in hydrogen bonding distance to an apparent iron-bound dioxygen molecule and sequence alignments show that this newly discovered active site Arg is conserved among all "Gln-type" CDO enzymes. Based on the position of this Arg we conclude that the "Gln-type" CDO homologs are not authentic CDOs.

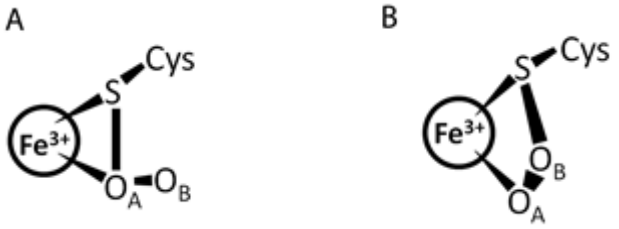
Introduction

In metazoa, fungi and prokaryotes the catabolism of cysteine begins with cysteine dioxygenase (CDO), a non-heme iron containing protein that converts cysteine to cysteine sulfinic acid (CSA) with the incorporation of both atoms of dioxygen according to the reaction:



The enzymatic mechanism of this enzyme is still not well understood despite recent enzymatic (1-5) and spectroscopic (6-9) characterizations of wild type and mutant CDOs, a number of crystal structure determinations for wild type CDO both without (10-12) and with (12-14) active site-bound ligands, studies with small molecule analogs of the metallocenter (15,16), and high level quantum mechanical analyses of potential mechanisms (17-20). All but one (5) of these studies have been done on either rat/mouse (identical in sequence) or human CDO (99% identical in sequence to rat CDO). As reviewed recently (21), for these mammalian CDOs the ferrous iron is coordinated by three His residues and an ordered water molecule, and nearby is a key Tyr157 residue (rat CDO numbering) that has a thioether crosslink with Cys93 and that is part of a Ser153-His155-Tyr157 catalytic triad. Cys, upon binding, displaces the Fe-bound water and ligates the iron via both its sulfhydryl and amine groups, with its carboxylate interacting with the side chains of Arg60, Tyr58 and the key catalytic Tyr157 (12).

For mammalian CDO, spectroscopic studies show that the oxygen binding site only becomes available after Cys binding (6), and crystal structures exist for a complex with only cysteine bound (PDB 4IEV; (12)) and for a complex with both Cys and oxygen bound and having reacted to form a Cys-persulfenate/persulfenic acid (Scheme 5.2A). The debate about the mechanism centers on whether cysteine thiolate is first oxidized by the iron proximal oxygen atom (to form a persulfenate like intermediate shown to the right) or by the iron distal oxygen atom to generate an FeIV-oxo intermediate (shown to the left):



This has been difficult to sort out due to the lack of spectroscopically discernable intermediates after Cys binding. In addressing questions of enzyme mechanism, it can be helpful to investigate divergent forms of an enzyme that are thought to conserve mechanism, because each provides a distinct window into the chemistry.

Among CDO enzymes, the bacterial homologs are quite different from the mammalian enzymes, with sequence identities in the 13-21% range (5). Of particular interest, these bacterial CDO homologs have a Gly residue in place of Cys93, and so do not have the Cys-Tyr cross-link present in mammalian CDOs. In addition, while many bacterial CDO homologs conserve the Arg60 residue that interacts with Cys α-carboxylate, some of them have a Gln residue at this position (5). Here we will call these "Arg-type" and "Gln-type" CDO homologs, respectively. One study of four "Arg-type" CDO homologs showed that they have CDO activity and can be thus considered CDOs. The only published characterization of a "Gln-type" CDO homolog, an enzyme from *Variovorax paradoxus*, showed that it had no CDO activity, but was active as a 3-mercaptopropionate dioxygenase (22). Structural studies of such bacterial CDO homologs would be valuable for the field, and even though they have not been described in the literature, two structures of such CDOs have been solved by structural genomics efforts and have their coordinates deposited in the Protein Data Bank (PDB entry 3EQE and 2GM6). One of these is a *Bacillus subtilis* enzyme that we refer to as *Bs*CDO, because it has been shown to convert Cys to CSA (5). It was solved at 2.8 Å resolution and is an Arg-type CDO homolog. The other is a functionally uncharacterized *Ralstonia eutropha* CDO homolog (*Re*CDOhom) that has been solved at 1.84 Å resolution and is a Gln-type enzyme. *Re*CDOhom is 55% identical in sequence to the *V. paradoxus* mercaptoproprionate dioxygenase leading us to suspect it will also have similar activity rather than being a CDO.

In the spirit of maximizing the value of the structures determined by structural genomics efforts, here we describe these structures and their comparisons with the mammalian CDO structures. In addition, we were able to reproduce the crystals of *Bs*CDO and solve the structure of its complex with cysteine in the active site. Highlights of this work are the highly conserved positions of the active site iron and of the Tyr157 equivalent residues despite the high divergence of bacterial CDO homologs, the observation of a diatomic molecule, modeled as O_2, bound to the active site iron in the *Re*CDOhom structure, and the discovery of an unanticipated active site Arg that is conserved among Gln-type bacterial CDO homologs.

Materials and Methods

*Bs*CDO expression and purification. A *Bs*CDO/pET32a expression plasmid was obtained from the DNASU Plasmid Repository (dnasu.asu.edu/DNASU/Home.jsp) in E. coli strain DH5α cells and used to transform E. coli strain BL21-DE3 chemically competent cells (Novagen). Our *Bs*CDO expression and purification basically follows the NESG protocols available in the PSI-knowledge base (23). Cells were grown at 37° C to OD600 ~0.6 in 3 L of Lysogeny Broth with 100 mg/mL ampicillin. The suspension was placed on ice and isopropyl-β-D-thiogalactopyranoside was added to be1 mM, before further growth at 17 °C to OD600 ~1.5 before harvesting. Lysis was carried out in 100mM Na2HPO4, 2.5 M NaCl, 100mM imidazole, pH 7.4 with protease inhibitors (20mM leupeptin, 1.5mM aprotinin, 1.5mM pepstatin and 100mM phenylmethylsulfonyl) and the clarified crude extract was incubated with 3 mL of cobalt resin (TALON®) at 4 °C for 1 h. The His-tagged *Bs*CDO was eluted with 20 mM Na2HPO4, 500 mM NaCl and 500 mM imidazole, pH=7.4 concentrated by Ultrafiltration (3K Filter Unit; Millipore) and run over a Superdex-75 gel filtration column (2.5 cm × 55 cm), equilibrated with 5 mM DTT, 100 mM NaCl, 20 mM Tris and 0.02% NaN_3 pH 7.5. Peak fractions were analyzed by SDS-PAGE, and the pure *Bs*CDO was pooled and concentrated to ~10 mg/mL, before being flash frozen in liquid N_2. Typical yield was ~2 mg of *Bs*CDO per liter of culture.

*Bs*CDO structure determination. Starting with crystallization conditions reported in the PDB entry header, crystals grew in 2 to 7 days at 298 K in hanging drops made of 4 uL of *Bs*CDO protein stock and 4 µL of a reservoir containing 18% (w/v) PEG 4000, 0.1 M potassium acetate, 0.05 M 2-(N-morpholino)ethanesulfonic acid (MES) at pH 6.0. For storage and soaks, crystals were transferred using a CryoLoopTM (Hampton Research) to artificial mother liquor (AML) identical to the reservoir solution. Crystals were harvested by pulling through a cryo-protectant solution made of an AML containing 20% glycerol and any relevant substrate/inhibitor before being plunged in liquid nitrogen. The cysteine complex crystal was soaked for 30 s with 100 mM cysteine at pH 7.0. We also attempted to get higher resolution data for unliganded *Bs*CDO crystals, but were not successful.

Data for Cys-bound *Bs*CDO were collected at the Advanced Light Source Lawrence Berkeley National Laboratory beam-line 5.0.1 in a gaseous nitrogen cryostream. Images were indexed and integrated with Mosflm (24) and scaled using Aimless (25). The crystal had space group $P4_322$ as previously reported for PDB entry 3EQE, and similar unit cell dimensions (Table 5.1). The high resolution cutoff criterion was that the recently introduced $CC_{1/2}$ statistic (26) should be ~0.2 (Table 5.1). Rf_{ree} flags were the same 5% of the reflections that were designated in PDB entry 3EQE.

*Bs*CDO model refinement at 2.3 Å resolution against the data from the Cys-soaked crystal began using PDB entry 3EQE with waters removed, and led to R/R_{free} values of 24.0/32.2%. Further refinements used Coot (27) for manual model building and Phenix (28) for minimizations. TLS refinement was done defining each chain as one TLS group. During iterative manual model building, sidechain rotomers were adjusted, Met1 was stubbed in chain B, an alternative conformation was added to Asn31 in chain B, and 69 water molecules were added in places having $2F_o$-F_c electron density ≥ 1 ρ_{rms}, F_o-F_c density ≥ 3 ρ_{rms}, and that were ≤ 3.7 Å from a potential hydrogen bonding partner. Ordered water molecules and or chloride atoms were insufficient to describe the active site density, as the waters' temperature factors

refined well below that of the iron atom and positive difference map features and unexplained $2F_o$-F_c density remained. Modeling a cysteine gave a nearly featureless difference map, and reasonable temperature factors and interactions with the active site residues. Molprobity (29) was used to monitor the stereochemical quality of the models during the final rounds of refinement. The final R/R_{free} at 2.3 Å resolution was 19.0/26.2% (Table 5.1).

The lower-resolution unliganded *Bs*CDO structure refinement against the deposited 2.8 Å resolution data from PDB entry 3EQE began from the refined Cys-complex model after Cys and active site waters were removed. Protocols were as above except for keeping all but 4 (which had shifted beyond 3.5 Å from the protein) of the waters even if they did not have $2F_o$-F_c electron density $\geq 1\ \rho_{rms}$, and the refinements quickly converged to R/Rf_{ree} values of 17.7/25.4 at 2.82 Å resolution (Table 5.1), an improvement over those values of 24.1/29.6% recorded for PDB entry 3EQE. In this final structure three density features near the iron that had been modeled as waters in 3EQE and in our refinement had B-factors near 40 $Å^2$, much lower than the B~70 $Å^2$ of the iron, suggesting that this was not the correct interpretation. Modeling it as a bound cysteine explained the electron density better, but the density features at this resolution were simply not sufficient for making this assignment and so we ended doing our final refinement with the three waters in the active site. However, so that the information about the possibility of this being a Cys-complex is not lost, we include in the PDB file the coordinates for a refined bound cysteine with their occupancy set to 0.0.

Polishing Refinement of *Re*CDOhom. Starting from the deposited coordinates and structure factors of *Re*CDOhom (PDB ID 2GM6), initial rounds of refinement in Phenix (30), using one TLS group and riding hydrogens, resulted in R/Rfree of 17.3/21.2% at 1.84 Å resolution. During manual model building, as described above for *Bs*CDO, additional water molecules, 5 additional ethylene glycols, and a sulphate were built. Alternative conformations were modeled for Glu38, Gly39 and 9 solvent molecules (ethylene glycol and water) and an alternate conformation was removed

from Leu124. The most significant change to the model was at the active site, where density that had been modeled as a sulphate was reinterpreted as a diatomic molecule, modeled as dioxygen. The R/R_{free} after this stage of refinement was 16.9/20.35% (Table 5.1), slightly below the 18.2/20.7% values of the deposited entry 2GM6.

At this point, original diffraction images for *Re*CDOhom provided by the JCSG were processed in the same way as the *Bs*CDO images and yielded data out to 1.65 Å resolution (Table 5.1). R_{free} flags were adopted from PDB entry 2GM6 out to its limiting resolution of 1.84 Å, and for the extended resolution range from 1.84 Å to 1.65 Å, a random 5% of the data were selected. At the extended resolution the maps were slightly better defined, allowing the solvent model to be refined further, and the active site dioxygen to be more precisely modeled. The final refined model had further improved R/R_{free} of 17.4/20.0 even at the extended 1.65 Å resolution.

Results and Discussion

Compared with the original PDB entries, the analysis of *Bs*CDO crystals with Cys-bound that diffract to 2.3 Å resolution and the reprocessing of the diffraction images from *Re*CDOhom to extend the resolutions from 1.84 to 1.65 Å have allowed us to generate improved models for both of these structures. The further refinements also led to lower overall R_{free} values by ~3 % for *Bs*CDO and ~1 % for *Re*CDOhom, even at the extended resolutions (Table 5.1). More importantly, the further refinements altered some key aspects of the models, most prominently the recognition that a diatomic O_2-like molecule rather than a sulfate is bound in the active site of *Re*CDOhom. Also, our work to reproduce the *Bs*CDO crystals led to the successful capture of a cysteine bound complex in its active site. In the rest of the Results and Discussion section we will first summarize broad features of the two proteins structures and then describe features seen in each active site and their implications.

Overall Structures. The structures of both the *B. subtilis* and *R. eutropha* enzymes show the expected cupin fold and overlay well with rat CDO despite the low overall sequence identities of ~21% and 18%, respectively; Figure 5.1A). The *Bs*CDO

structures have two copies in the asymmetric unit with the chains agreeing within ~0.5 Å for both the unsoaked and Cys-bound structures. Chain A includes residues 1-154 and Chain B residues 1-152 (missing 7 and 9 C-terminal residues, respectively). One difference between the chains is that due to crystal contacts, a loop at residues 98-102 and the C-terminus of Chain A are more ordered. We base our descriptions here on chain A, but they also match Chain B unless noted. For *Re*CDOhom, there is only one chain in the asymmetric unit and the model includes residues 11-202, with all parts of the chain having reasonably well-defined electron density. As can be seen in a structure-based sequence alignment of the two bacterial CDO homologs along with rat CDO (Figure 5.1B), the amino acid sequence of *Bs*CDO is on the whole ~40 residues shorter than *Re*CDOhom and rat CDO. Most of this (~20 residues) is due to a shorter C-terminus and the ordered backbone of *Bs*CDO stops before strand β11 even though there enough residues present in the sequence to form it (Figure 5.1). In terms of changes in secondary structure, the one other notable feature is that relative to rat and *Bs*CDO, *Re*CDOhom has a β-bulge insertion in the middle of strand β6 (Figure 5.1).

Active Sites

*Bs*CDO active site. We will first describe the 2.3 Å resolution structure with Cys-bound since that is the more informative structure. Similar to previous structurally characterized CDO enzymes (11-14), the active site of un-soaked *Bs*CDO has density for a non-heme iron that is coordinated by three conserved residues His75, His77, His125 (Figure 5.2A). Also seen in the active site are well-ordered side chains of key residues Tyr141, Arg50 and Tyr48 (corresponding to Tyr157, Arg60 and Tyr58 in rat CDO). Nearby are Ser137 and His139 that with Tyr141 form the catalytic triad, and as expected no crosslink is formed with Tyr141. The Gly residue substituting for the Cys93 of mammalian CDOs is well-ordered. A unique aspect of *Bs*CDO is that Ser137 of the catalytic triad does not receive a hydrogen bond as the Trp77 seen in mammalian CDO (13) is replaced with Ile66 in *Bs*CDO and no other residue is hydrogen bonding distance to Ser137. In the Cys-soaked crystal, additional strong active site $2F_o$-F_c density present was well fit by cysteine (Figure 5.2A). Because of

the medium resolution (2.3 Å) of the structure, we waited to model the cysteine until the final rounds of refinement, only adding it after eliminating plausible alternative interpretations of the density (see methods). In this position, Cys coordinates the iron in a bi-dentate fashion via its Sγ and N atoms, with the α-amino group location being well-defined stereochemically even though it is not well defined by electron density at this resolution (Figure 5.2A). The α-carboxylate makes well-aligned hydrogen-bonds with Arg50, Tyr48 and the Cys α-amino group.

For the 2.8 Å resolution structure of the unsoaked crystal (based on the diffraction data collected by NESG), the iron and active site residues are positioned very similarly to how they are seen in the Cys-complex (Figure 5.2B). As was done for PDB entry 3EQE, we attempted to model three peaks around the iron as water molecules. However, in such a model the waters refined to have temperature factors of ~35 $Å^2$, which are much lower than the values of ~50-60 $Å^2$ seen for the surrounding side chains and the iron atom. This implies that these peaks are actually not due to water. Noting that the density peaks match reasonably well with the peaks of density seen in the 2.3 Å Cys complex, we suggest that the density actually is due to a bound cysteine (or similar molecule) (see Figure 5.2B – semitransparent model). Refining a cysteine in the structure led to reasonable B-factors and a clean difference map, supporting this assignment. Nevertheless, because at this resolution we cannot definitively interpret the active site density and it may be that Cys is not bound, to be conservative, our final refinement and deposition used the minimal interpretation having the three waters in the active site; to make sure users of the coordinates are aware of the possibility of Cys being bound, we also include in the coordinate file, with occupancy set to zero, the coordinates and refined temperature factors we obtained by refining a bound Cys at 100% occupancy.

<u>*Re*CDOhom active site.</u> Unlike the four-coordinate iron seen in ligand-free mammalian CDO, the *Re*CDOhom crystal structure has a metal that is six-coordinate with the coordination sphere including three histidines (His94, His96, His147), two water (or hydroxide) molecules and a diatomic molecule we have modeled as

dioxygen (Figure 5.2C). Consistent with it being a dioxygen, the diatomic molecule makes a 123˚ angle in its approach to the iron, close to the 120˚ angle expected for an sp2 hybridized atom, and also is accepts a hydrogen bond from Arg173 to the iron-proximal atom. Furthermore, when modeled as dioxygen, there are no residual difference map features and the B-factors refine to values of ~40 $Å^2$ that are comparable to the water ligands (~40 $Å^2$) and other surrounding atoms (~34 $Å^2$ for the Fe, 50 $Å^2$ for Tyr164 and 35 $Å^2$ for Arg173). For these reasons we will refer to it as a dioxygen, while recognizing it could be something else. With regard to other active site residues, the Ser160-His162-Tyr164 catalytic triad is conserved, and like the mammalian CDOs (13), this enzyme has a Trp (Trp85) that hydrogen bonds with Ser160. The Tyr164 hydroxyl is close (~3.0 Å) to the iron-bound oxygen, but we consider this a van der Waals interaction, as based on geometry Tyr164 donates a hydrogen bond to the iron bound water/hydroxide (Figure 5.2C). Again, no Cys-Tyr crosslink is present, and the Gly replacing Cys93 is well-ordered.

As was noted in the introduction, the typical active site Arg is replaced with Gln67, and in this structure the Gln67 side chain is only weakly ordered. In contrast, the side-chain of Arg173 is well ordered, with its guanidino group sandwiched between the side chains of Ile168 and Phe185; in addition to its interaction with dioxygen, it makes hydrogen bonds with Asp95-O, Ser187-OH, and a well-ordered water molecule that in turn hydrogen bonds with backbone atoms Val171-O and Ser187-NH. Also, related to the position of Arg173 are the presence in *Re*CDOhom relative to rat CDO and *Bs*CDO of a single residue deletion after strand β3 (at Thr97) and a two residue insertion after strand β9 making a 3_{10}-helix (including Ile168) (Figure 5.1B). These two segments pack against each other in a way that provides the environment of Arg173.

Comparison of *Bs*CDO, *Re*CDOhom and rat CDO active sites. In figure 5.2D, we present an overlay of the *Bs*CDO active site with Cys bound, the *Re*CDOhom active site with dioxygen bound, and the previously published (12) rat CDO active site with Cys-persulfenate bound. This overlay shows the remarkably consistent relative

placement of the iron and the residues equivalent to rat CDO Tyr157 and Tyr58 in all three structures. Also visible in *Bs*CDO and rat CDO are their common binding mode of the Cys and the placement of the residue equivalent to rat CDO Arg60. This provides quite strong evidence that *Bs*CDO and other bacterial CDOs conserving Arg60 will be mechanistically highly similar to rat CDO despite the lack of the Cys-Tyr crosslink.

Also of particular interest is what can be concluded about *Re*CDOhom and related enzymes that have a Gln in the place of Arg60. The side chain Gln67 in *Re*CDOhom is located similarly to that of the Arg in rat and *Bs*CDO, but would not be able to conserve all of interactions that Arg makes with the α-carboxylate of Cys. Arg173 of *Re*CDOhom could of interest as a potential replacement for interacting with the Cys carboxylate, but it also is not suitably positioned to play that role. Also, rather surprisingly, the dioxygen binding site matches the position filled in rat and *Bs*CDOs by the α-amino group of Cys, rather than the position of oxygen binding in CDOs as is inferred from crystal structures (12-14) (Figure 5.2D), spectroscopic results (6,7) (9) and calculations (17,18). An important question is whether these observations imply that the dioxygen site seen is a non-productive binding mode or perhaps reflects that the Gln-type CDO homologs bind oxygen differently. While not able to answer this question in the absence of a structure of a complex with a productive substrate bound, we can already conclusively say that this enzyme cannot bind Cys in the same way as do the CDOs. This is because Arg173, as an obligatory hydrogen bond donor, would form an unfavorable clash with the hydrogen of a bound α-amino group that, because its lone pair of electrons are interacting with the iron, would also be an obligatory hydrogen bond donor in that interaction.

Arg173 as a key residue for Gln-type CDO homologs

A structure based sequence alignment between *Bs*CDO, *Re*CDOhom and rat CDO provides a robust alignment for examining equivalent residues and illustrates minor variations in the secondary structure between these divergent CDO homologs (Figure 5.1B). The new Arg173 active site residue of *Re*CDOhom aligns with a Met

and a Cys in the two CDOs, but is also conserved in identity and position in the structure of another unpublished Gln-type CDO homolog (PDB ID 3USS). To assess if this Arg173 is conserved among Gln-type CDO enzymes, the residue conservation pattern among many CDO homologs was generated by combining the unique sequences from separate blast searches done using either *Bs*CDO, *Re*CDOhom and rat CDO. This group of sequences shows only a low level of conservation of Arg173 (Figure 5.3A).

However, splitting these sequences into Arg-type and Gln-type groups is more informative. At the position of Arg173 of *Re*CDOhom, the Arg-type sequences tend to have a nonpolar residue, but without strong conservation (Figure 5.3B). In contrast, the Gln-type sequences conserve the Arg173 position rather well (Figure 5.3C), and also conserve the single residue deletion after strand β3 and the two residue insertion after strand β9. This conservation is consistent with a key functional role for Arg173 in the Gln-type CDOs, and also with the conclusion that no one of the Gln-type enzymes are authentic CDOs – since they could not bind the Cys α-amino group in the expected way. Taken into account the observation that the *V. paradoxus* 3-mercaptoproprionate dioxygenase is a Gln-type enzyme with 55% sequence identity with *Re*CDOhom, we propose that all of the Gln-type enzymes with be dioxygenases with specificity for a substrate activity that the putative oxygen molecule interacting with Arg173 in the active site of *Re*CDOhom. We further suggest that given a structure available for *Re*CDOhom that is of high quality (1.8 Å resolution) and is a potentially very informative complex with a molecular oxygen bound, *Re*CDOhom would be an excellent system for pursuing kinetics and spectroscopic studies of catalysis that would bring new insights into this family of dioxygenases.

Making the Most of Structural Genomics Structures

The NIH funded Protein Structure Initiative (PSI) invested heavily in structural genomics research centers with the dual goals of solving the structure of many representative proteins and protein domains as well as developing high throughput structure determination techniques (23). These efforts were fruitful on both fronts and

have accounted for more than 13,100 solved structures (PDB query 06/2014). One unforeseen consequence of these efforts is that there now exist thousands of entries in the PDB for protein structures that have not been described in the peer-reviewed literature. As is, these structures are of limited value to the broader scientific community both because they will not show up in literature searches and because no expert having both knowledge of protein crystallography and the particular protein family has carefully vetted the structures for accuracy and for information content. We suggest that, as exemplified by our work here on CDO homologs and elsewhere for two sets of peroxiredoxin structures (31,32), there exists now a rich opportunity for researchers with appropriate expertise to make more accessible the many unpublished fruits of structural genomics that are ripe but as yet unharvested for general consumption.

Acknowledgements

We thank Rick Cooley and Dale Tronrud for useful discussions and Ryan Mehl for the use of his protein purification facilities. We also thank Ashley Deacon and the Joint Center for Structural Genomics for providing us with the original images collected from *Re*CDOhom crystals. This project was supported in part by Grant DK-056649 to PAK and Martha H. Stipanuk (Cornell) from the National Institute of Diabetes and Digestive and Kidney Diseases. Synchrotron data were collected at the Advanced Light Source, supported by contract DE-AC02-98CH10886 from the Office of Basic Energy Sciences of the U.S. Department of Energy.

Accession Numbers

Coordinates and structure factors for the *Bs*CDO and *Re*CDOhom models have been deposited in the Protein Data Bank (lower-resolution *Bs*CDO (PDB code 4QM8); Cysteine-bound *Bs*CDO (PDB code 4QM9); *Re*CDOhom (PDB code 4QMA)).

Table 5.1. Data collection and Refinement Statsistics for *Bs*CDO and *Re*CDOhom

	***Bs*CDO unsoaked** [a]	***Bs*CDO Cys-soak**	***Re*CDOhom**
Data collection			
Unit Cell (Å)	a=b=65.8, c=197.3	a=b=65.5, c=199.4	a=b=57.01, c=216.70
Resolution (Å)	50-2.82 (2.95-2.82)	50-2.30 (2.30-2.38)	45-1.65 (1.68-1.65)
Unique Obs.	19930 (2657)	20271 (1939)	44377 (2185)
Multiplicity	14.2 (9.8)	23.4 (13.7)	21.0 (16.5)
Completeness	98.9 (92.3)	100.0 (100.0)	100 (100)
Average I/σ	– (3.6)	13.4 (1.0)	14.1 (0.6)
R_{meas} (%)[b]	0.086 (0.63)	0.186 (2.89)	0.113 (5.33)
$CC_{1/2}$ (%)[c]	–	0.999 (0.25)	1.0 (0.28)
Res <I/σ>~2 (Å)[c]	–	2.5	1.85
Refinement			
R_{cryst} / R_{free} (%)	17.7/25.4	19.0/26.2	17.4/20.0
No. residues	308	308	192
No. waters	63	67	209
No. atoms	2422	2437	3342
rmsd angles (°)	1.174	1.060	1.213
rmsd lengths (Å)	0.011	0.012	0.014
φ,ψ favored (%)[d]	94	95	96
φ,ψ outliers (%)[d]	0	0.33	0
<B> protein ($Å^2$)	69	61	48
<B> Fe ($Å^2$)	63	47	38
<B> Cys ($Å^2$)	83[e]	64	
<B> O_2 ($Å^2$)			49
<B> solvent ($Å^2$)[f]	63	57	57
PDB code	4QM8	4QM9	4QMA

[a] Data collection statistics as reported in the original PDB entry 3EQE

[b] R_{meas} is the multiplicity-weighted merging *R* (33). For 4QM8, R_{merge} is reported.

[c] $CC_{1/2}$ is the correlation between two half datasets as defined in Karplus & Diederichs (*25*). Resolution at which <I/σ>~2 for comparison with previous high resolution cutoff criteria (26).

[d] Ramachandran statistics as defined by Molprobity (29).

[e] For Cys at full occupancy; included in the deposited structure with occupancy=0 to reflect uncertainty in the interpretation.

[f] Solvent in *Bs*CDO are waters, and in *Re*CDOhom are water, ethylene glycol and sulfate.

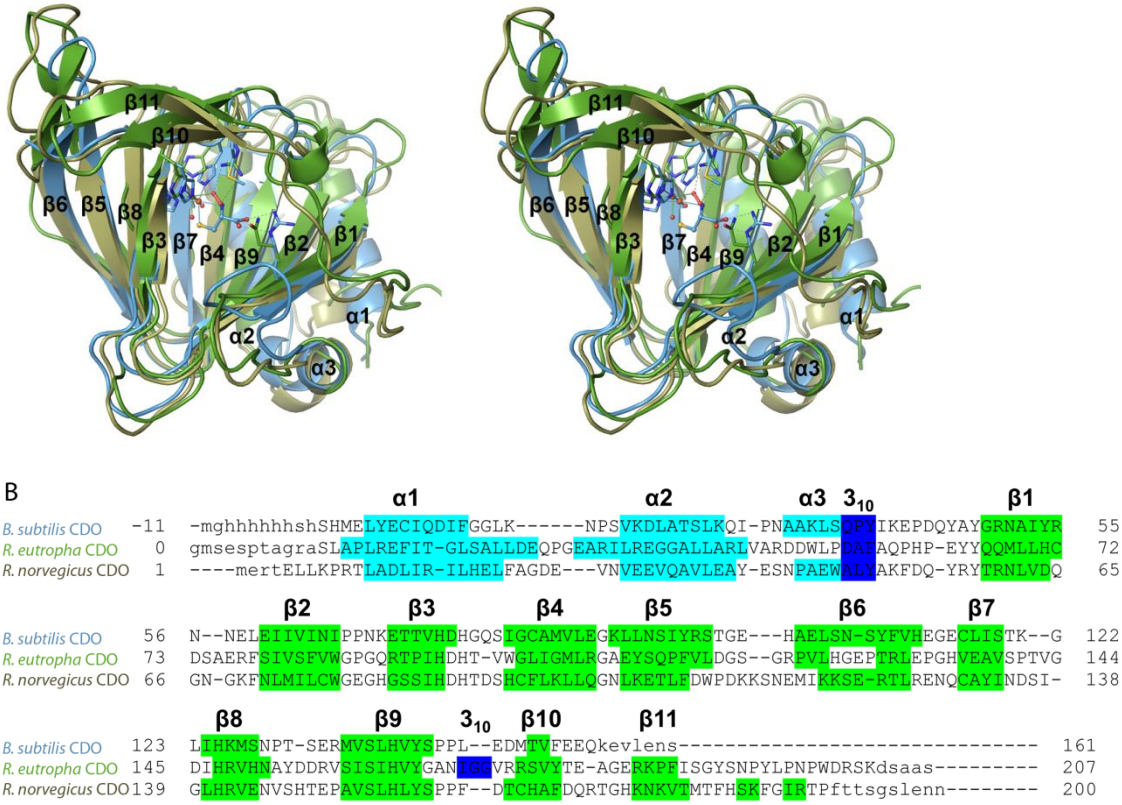


Figure 5.1. Common cupin-fold of the bacterial CDO homologs. (A) Stereoview of the overlaid ribbon diagrams of *R. norvegicus* CDO (gold; PDB 4IEU); (12)), *Bs*CDO (blue; ~2.3 Å Cα-rmsd vs. rat CDO) and *Re*CDOhom (green; ~2.2 Å Cα-rmsd vs. rat CDO) shows their similar overall structures. The active site Arg/Gln and the new active site Arg of the *Re*CDOhom (Met in rat CDO) are shown along with the iron coordinating His residues and the cysteine substrate as bound to rat CDO and *Bs*CDO. The secondary structure labels are shown, with all three alpha helicies being on the N-terminal side of the beta sheets. The overlay was generated using CEalign implemented in Pymol (34). (B) The structure based sequence alignment of *Bs*CDO *Re*CDO homolog, and *R. norvegicus* CDO as generated using PROMALS (35), and manually colored according to secondary structure as defined by DSSP (36).

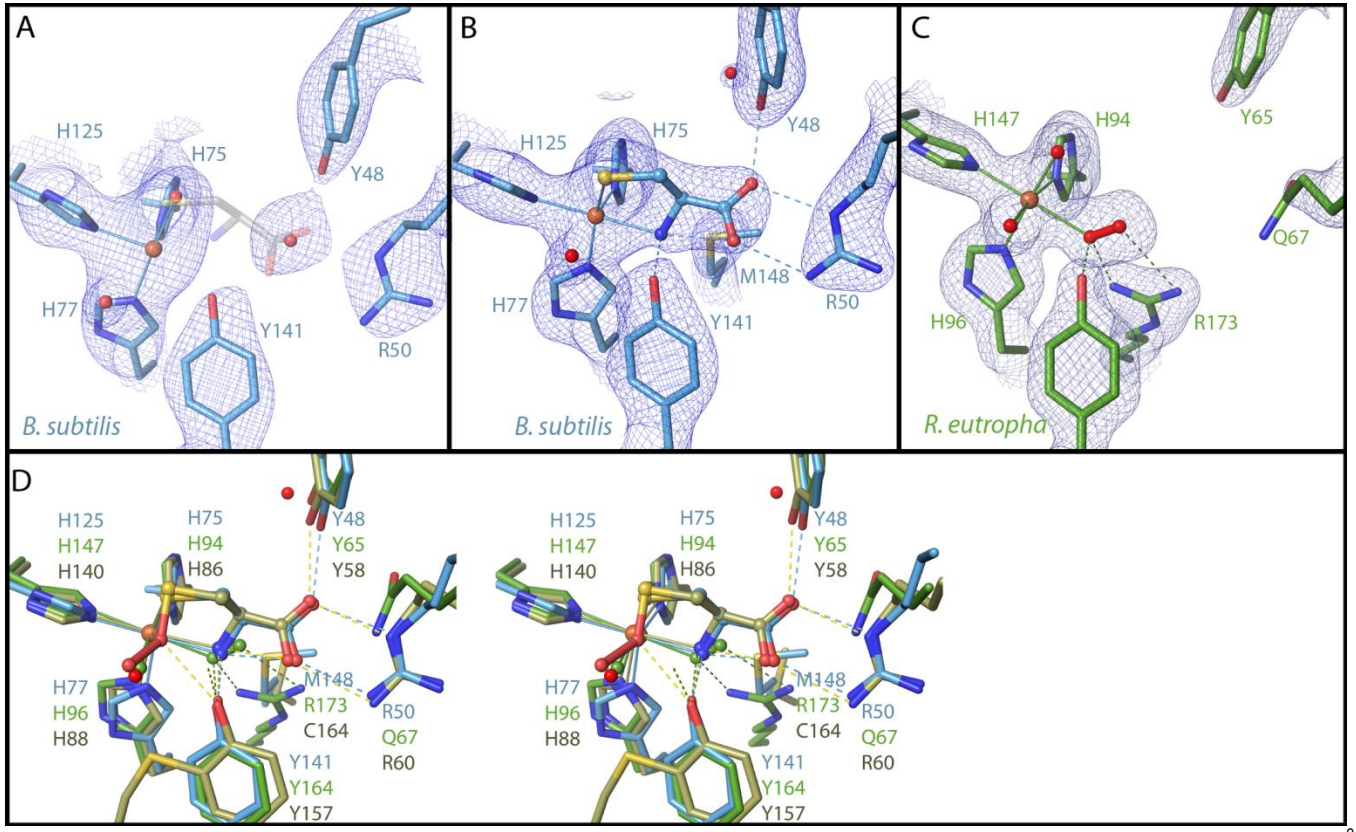


Figure 5.2. Active site structures of *Bs*CDO and *Re*CDOhom. Active site density for (A) *Bs*CDO at 2.3 Å resolution with cysteine-bound, (B) unsoaked *Bs*CDO at 2.8 Å resolution, and (C) *Re*CDOhom at 1.65 Å resolution with a diatomic molecule bound to the metal. All maps are $2F_o$-F_c density contoured to ~ 1.2 ρ_{rms}. (D) Local overlay of the active sites of *Bs*CDO (blue carbons), rat CDO (gold carbons) (12), and *Re*CDOhom (green carbons).

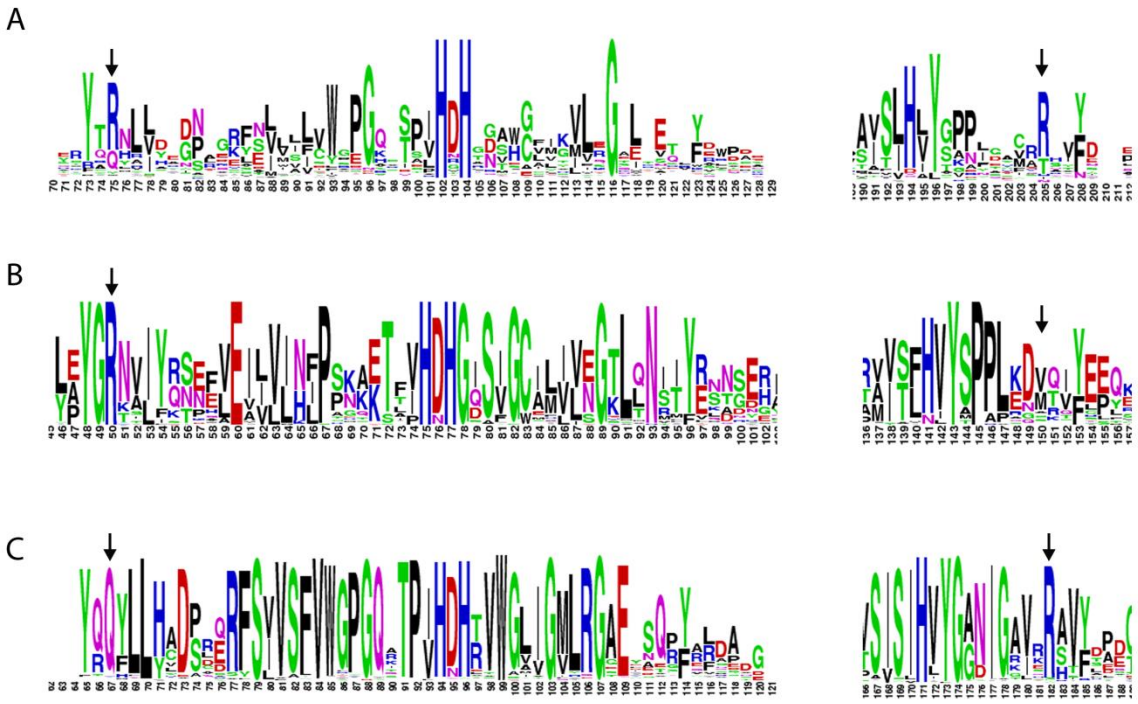


Figure 5.3. Residue conservation patterns among relevant CDO homologs. (A) Conservation pattern among CDO sequences obtained by combing 717 unique sequences from three PROSMALS PSI-blast searches using *Bs*CDO, rat CDO, and *Re*CDOhom as templates against the UNIPROT 90 database with sequences included down to a 20% sequence identity threshold. (B) Arg-type subset residue conservation pattern. Sequences were obtained by a BLASTP search of *Bs*CDO, with all 101 sequences above an e^{-16} threshold included in an alignment. (C) Gln-type CDO homolog subset residue conservation pattern. Sequences were obtained by a BLASTP search of *Re*CDOhom, with all 151 sequences above an e^{-16} threshold included in the alignment. Conservation pattern images were generated using WebLogo (37).

References

1. Dominy, J. E., Jr., Hwang, J., Guo, S., Hirschberger, L. L., Zhang, S., and Stipanuk, M. H. (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. J Biol Chem 283, 12188-12201

2. Ye, S., Wu, X., Wei, L., Tang, D., Sun, P., Bartlam, M., and Rao, Z. (2007) An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. J Biol Chem 282, 3391-3402

3. Siakkou, E., Rutledge, M. T., Wilbanks, S. M., and Jameson, G. N. (2011) Capturing crosslink formation with enzymatic activity in cysteine dioxygenase. Biochim Biophys Acta

4. Imsand, E. M., Njeri, C. W., and Ellis, H. R. (2012) Addition of an external electron donor to in vitro assays of cysteine dioxygenase precludes the need for exogenous iron. Arch Biochem Biophys 521, 10-17

5. Dominy, J. E., Jr., Simmons, C. R., Karplus, P. A., Gehring, A. M., and Stipanuk, M. H. (2006) Identification and characterization of bacterial cysteine dioxygenases: a new route of cysteine degradation for eubacteria. J Bacteriol 188, 5561-5569

6. Pierce, B. S., Gardner, J. D., Bailey, L. J., Brunold, T. C., and Fox, B. G. (2007) Characterization of the nitrosyl adduct of substrate-bound mouse cysteine dioxygenase by electron paramagnetic resonance: electronic structure of the active site and mechanistic implications. Biochemistry 46, 8569-8578

7. Crawford, J. A., Li, W., and Pierce, B. S. (2011) Single turnover of substrate-bound ferric cysteine dioxygenase with superoxide anion: enzymatic reactivation, product formation, and a transient intermediate. Biochemistry 50, 10241-10253

8. Gardner, J. D., Pierce, B. S., Fox, B. G., and Brunold, T. C. (2010) Spectroscopic and computational characterization of substrate-bound mouse cysteine dioxygenase: nature of the ferrous and ferric cysteine adducts and mechanistic implications. Biochemistry 49, 6033-6041

9. Li, W., Blaesi, E. J., Pecore, M. D., Crowell, J. K., and Pierce, B. S. (2013) Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O(2) coupling efficiency in mouse cysteine dioxygenase. Biochemistry 52, 9104-9119

10. McCoy, J. G., Bailey, L. J., Bitto, E., Bingman, C. A., Aceti, D. J., Fox, B. G., and Phillips, G. N., Jr. (2006) Structure and mechanism of mouse cysteine dioxygenase. Proc Natl Acad Sci U S A 103, 3084-3089

11. Simmons, C. R., Liu, Q., Huang, Q., Hao, Q., Begley, T. P., Karplus, P. A., and Stipanuk, M. H. (2006) Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. J Biol Chem 281, 18723-18733

12. Driggers, C. M., Cooley, R. B., Sankaran, B., Hirschberger, L. L., Stipanuk, M. H., and Karplus, P. A. (2013) Cysteine dioxygenase structures from pH4 to 9: consistent cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH. J Mol Biol 425, 3121-3136

13. Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Jr., Begley, T. P., Stipanuk, M. H., and Karplus, P. A. (2008) A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. Biochemistry 47, 11390-11392

14. Souness, R. J., Kleffmann, T., Tchesnokov, E. P., Wilbanks, S. M., Jameson, G. B., and Jameson, G. N. (2013) Mechanistic implications of persulfenate and persulfide binding in the active site of cysteine dioxygenase. Biochemistry 52, 7606-7617

15. McQuilken, A. C., Jiang, Y., Siegler, M. A., and Goldberg, D. P. (2012) Addition of dioxygen to an N4S(thiolate) iron(II) cysteine dioxygenase model gives a structurally characterized sulfinato-iron(II) complex. J Am Chem Soc 134, 8758-8761

16. McQuilken, A. C., and Goldberg, D. P. (2012) Sulfur oxygenation in biomimetic non-heme iron-thiolate complexes. Dalton Trans 41, 10883-10899

17. Kumar, D., Thiel, W., and de Visser, S. P. (2011) Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. J Am Chem Soc 133, 3869-3882

18. Kumar, D., Sastry, G. N., Goldberg, D. P., and de Visser, S. P. (2012) Mechanism of S-oxygenation by a cysteine dioxygenase model complex. J Phys Chem A 116, 582-591

19. Aluri, S., and de Visser, S. P. (2007) The mechanism of cysteine oxygenation by cysteine dioxygenase enzymes. J Am Chem Soc 129, 14846-14847

20. de Visser, S. P., and Straganz, G. D. (2009) Why do cysteine dioxygenase enzymes contain a 3-His ligand motif rather than a 2His/1Asp motif like most nonheme dioxygenases? J Phys Chem A 113, 1835-1846

21. Stipanuk, M. H., Simmons, C. R., Karplus, P. A., and Dominy, J. E., Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. Amino Acids 41, 91-102

22. Bruland, N., Wubbeler, J. H., and Steinbuchel, A. (2009) 3-mercaptopropionate dioxygenase, a cysteine dioxygenase homologue, catalyzes the initial step of 3-mercaptopropionate catabolism in the 3,3-thiodipropionic acid-degrading bacterium *variovorax paradoxus*. J Biol Chem 284, 660-672

23. Berman, H. M., Westbrook, J. D., Gabanyi, M. J., Tao, W., Shah, R., Kouranov, A., Schwede, T., Arnold, K., Kiefer, F., Bordoli, L., Kopp, J., Podvinec, M., Adams, P. D., Carter, L. G., Minor, W., Nair, R., and Baer, J. L. (2009) The

protein structure initiative structural genomics knowledgebase. Nucleic Acids Research 37, D365-D368

24. Leslie, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EAMCB newsletter on protein crystallography 26

25. Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr D 62, 72-82

26. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. Science 336, 1030-1033

27. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D 66, 486-501

28. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D 66, 213-221

29. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D 66, 12-21

30. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213-221

31. Gretes, M. C., and Karplus, P. A. (2013) Observed octameric assembly of a Plasmodium yoelii peroxiredoxin can be explained by the replacement of native "ball-and-socket" interacting residues by an affinity tag. Protein Sci 22, 1445-1452

32. Perkins, A., Gretes, M. C., Nelson, K. J., Poole, L. B., and Karplus, P. A. (2012) Mapping the active site helix-to-strand conversion of CxxxxC peroxiredoxin Q enzymes. Biochemistry 51, 7638-7650

33. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4, 269-275

34. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11, 739-747

35. Pei, J., Kim, B. H., Tang, M., and Grishin, N. V. (2007) PROMALS web server for accurate multiple protein sequence alignments. Nucleic Acids Res 35, W649-652

36. Joosten, R. P., te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs. Nucleic Acids Res 39, D411-419

37. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. Genome Res 14, 1188-1190

Chapter 6

Concluding Discussion and Outlook

Summary

In this dissertation, I have described structural studies I carried out on two enzymes involved in sulfur chemistry. In addition to the specific contributions to understanding SsuE and CDO chemistry, the studies reported here demonstrate the value that structural approaches have for advancing our understanding of biology. The insights provided have largely been based on the determination of X-ray crystal structures, in that most if not all of the major conclusions were possible because structural considerations guided complementary efforts such as phylogenetic analysis, spectroscopic experiments, analytical ultracentrifugation experiments, and reconsiderations of interpretations of published kinetics data. Each of the major contributions described in this dissertation is an example of the power that structure function studies, combined with the concepts of evolution and data from functional studies, have to provide insight in biological systems.

In this chapter I present conclusions of the SsuE work, and then address the CDO project. As I see it, SsuE is a more complete study in terms of the level of understanding we have developed, because although we have made great strides towards understanding CDO, there are still many mysteries to be solved even about its basic enzymatic mechanism. For SsuE, I will present the main conclusions, anticipated impacts and suggest some logical future studies. Then for CDO, I will similarly present the main conclusions and anticipated impacts, and speculate on future studies that might shed light on some of the remaining mysteries. Finally, I will finish the dissertation with some concluding remarks.

SsuE structure-function studies

Main conclusions and anticipated impacts

SsuE has an FMN-leading mechanism. My work on SsuE has provided a major advance to our understanding of SsuE chemistry and not only provided substantial clarifying insight into how SsuE works, but also into the functioning of members of

the broader flavodoxin-like superfamily and two-component monooxygenase systems. In my mind, the most important discovery from our SsuE structural work was that an FMN-leading mechanism is occurring with FMN binding in the first site before the second site is made available. I anticipate that the biggest long term impacts of this work will be from the general reaction cycle that emerged and how this emphasizes that related reductases can be following the same fundamental general mechanism even when there may be notable differences in the details of their properties. I am hopeful that because we have now resolved the discrepancy between the reported mechanisms of SsuE and EmoB, the two-component field will no longer group FMN-reductases by whether they purify with flavin bound or not (1), but will embrace grouping enzymes by their evolutionary relationships. However, I have also learned that deeply seated views can be slow to change. As emphasized by our ability to identify and correct the mistaken conclusions made based in the earlier kinetics studies, grouping by evolution will substantially help the two-component field. In particular, for those scientists studying FMN-reductases from the flavodoxin-like superfamily, the general reaction cycle can be used as a framework to guide both the design and the interpretation of future mechanistic studies.

SsuE's quaternary structure and weak binding to $FMNH_2$. Two other main conclusions specifically advance our understanding of SsuE. The first is the structure of the SsuE tetramer and dimer and the discovery that it is an obligate dimer. Additionally, that FMN binding favors the dimer, unlike those related FMN-reductases that lack a π-helix, has implications for SsuE/SsuD kinetics and associations, as the dimer/tetramer equilibrium could influence the rate of the overall reaction by influencing the association between SsuE and SsuD, particularly by influencing the rate of FMN transfer. Future work to better understand the dimer/tetramer equilibrium, especially in the presence of SsuD, will be valuable. The second conclusion that advances our understanding of SsuE is our proposal for the origins for the lower affinity of SsuE for $FMNH_2$ compared to FMN, where the hydrogen from the reduced flavin clashes with a backbone amide hydrogen (Figure

6.1). Flavodoxins undergo a peptide flip to accommodate $FMNH_2$ and have a Gly at this position. The inability of SsuE to undergo the peptide flip is one reason SsuE will have a tendency to follow the inner circle of the reaction cycle (Figure 2.8). Of these two possible routes, we do not know which occurs *in vivo* or even if both paths can contribute to productive $FMNH_2$ transfer to the partner monooxygenase.

Future studies

How does varying FMN concentration influence specific properties of SsuE? To test and further explore our hypothesis that its specific properties and environment influence which route around the reaction cycle SsuE takes, it would be very informative to study how varying the FMN concentration over a much wider range than has been done changes the measured kinetics as well as the dimer-tetramer equilibrium. Also providing an intriguing opportunity for a follow up study would be seeking to control the path around the general catalytic cycle by modulating the peptide flip. For this purpose, a mutation of Lys77 to Gly, if it allowed a peptide flip and did not cause other problems, should greatly increase SsuE's affinity for $FMNH_2$. Structural studies could be carried out on this K77G variant to test if it can undergo a peptide flip upon FMN reduction. If so, this K77G variant would have the first FMN stay bound during the catalytic cycle and allow for the test to see if the second $FMNH_2$ can be transferred to SsuD without having to greatly increase FMN concentration. Additionally, if the in vitro studies indicated the mutant had the desired properties, mutation of the gene coding for SsuE on the chromosome of *E. coli* such that it encoded the K77G variant could be used to test if the two-component system was more or less functional *in vivo* when SsuE was forced around the outer path of the reaction cycle.

How does flavin transfer occur? The mechanism of flavin transfer to SsuD is also not understood (2,3). Ideally a crystal structure of an SsuE/SsuD complex would provide the greatest insight, but as there are just over 200 heterocomplexes in the PDB, this is a difficult task. However, the structure of SsuD is known and so with the

structure of SsuE now in hand SsuE/SsuD docking studies combined with mutation studies could be used to elucidate the epitope and the role that direct interactions play in the overall sulfur acquisition chemistry of the SsuE/SsuD system. Of particular interest would be to delete Tyr118, turning the π-helix back into an α-helix to investigate the role that the π-helix Tyr plays in the Kd for FMN and the kinetics of the reaction as well as its role in SsuD association and quaternary structure.

Studies to advance our understanding of SsuE's close homologs. Future studies aimed at providing further insight into SsuE's closest homologs, EmoB and the uncharacterized 3k1y would also be very nice. Specifically, knowing the Kd for $FMNH_2$ would allow us to better understand if EmoB is more likely to follow the outside route than SsuE or if it was simply due to the conditions of the kinetics study. A full characterization of 3k1y would likely confirm that it behaves similarly to SsuE and EmoB.

CDO structure-function studies

Main conclusions and anticipated impacts

The emerging role of cysteine dioxygenase in cancer progression and detection as well as its role in Alzheimers and other diseases has revealed that understanding CDO is not just important for understanding sulfur chemistry but also, potentially, for improving human health. Although much remains to be discovered, we have made significant advances in understanding CDO chemistry and have provided a solid structural foundation on which to build future work. I will highlight what I think are the most important conclusions and their impacts.

*The persulfenate forms in the active site from pH 5.5 to 7.0***.** Our most important result from the exhaustive pH dependence study was showing that persulfenate is captured in the active site of CDO from pH 5.5 to 7.0, and taken together with recent results showing that it is formed in the active site (4), we have

successfully advanced the field by proving that the Cys-persulfenate is energetically accessible in the active site of CDO from pH 5.5 to 7.0. Although, in itself, this does not show if the Cys-persulfenate is on-pathway or off-pathway, it does provide a solid basis for future studies.

Our high-resolution Cys-bound complex is the first reliable Cys-only complex. The fortuitous capturing of Cys in the active site of CDO is important for impacting future studies because it provides the first reliable complex of Cys bound in the active site. As discussed in Chapter 3, the previously reported Cys-bound model (PDB entry 2ic1) did not actually have Cys bound. Because much of the work towards understanding CDO chemistry is carried out using small molecules and quantum mechanical calculations as tools, having a reliable high resolution Cys-bound complex is of the utmost importance.

The role of the crosslink and the catalytic Tyr in Cys binding and catalysis. The critical role of the crosslink in the active site is apparent even when examining just the unliganded active site density and observing how the Tyr side chain shifts such that it is no longer well positioned to participate in the catalytic triad. The importance of having the precise Tyr location is underscored even in these first structures by the observed dramatic influence on the iron coordination, as a water is exchanged for a chloride. However, the true extent of the disruption was seen when soaking Cys into the active site. Unexpectedly, the Tyr is important for modulating the pKa of the Cys α-amino group, as demonstrated by the fact that the crosslinked form is able to bind Cys/Cys-persulfenate from pH 5.5 to 9 via both its thiolate and α-amino groups, but that the non crosslinked form binds to the iron only though the thiolate. These thiolate-only binding modes are not productive for catalysis, as the sulfur partially blocks the oxygen binding site inferred from the Cys-persulfenate studies. Given the amount of disruption that we see in the uncrosslinked active site, it is not surprising that no persulfenate formation occurs.

The mechanism of homocysteine inhibition. Homocysteine inhibition may be relevant to disease, and so defining a mechanism for homocysteine inhibition was an important contribution of this work. Homocysteine binds to the active site iron via only its thiol, blocking both the oxygen binding site and the Cys binding site in non crosslinked and crosslinked CDO. With homocysteine being able to bind to all forms of CDO it would not only inhibit CSA production but would also inhibit crosslink formation.

Bacterial "Arg-type" CDOs bind Cys similarily and "Gln-type" homologs are not CDOs. By obtaining a well-resolved structure of *Bs*CDO in the presence of Cys, we were able to obtain a Cys bound complex for the "Arg-type" bacterial CDOs and show that Cys binds similarly as it does in mammalian CDO. This structure also revealed that the catalytic triad is intact, that the Tyr is not shifted relative to mammalian CDO and that the active site looks more like the active site in crosslinked mammalian CDO than it does the non crosslinked variants. Taken together, these results make *Bs*CDO an interesting system to use in future studies since it is so similar to mammalian CDO but does not form the crosslink. "Gln-type" CDO homologs have a very different active site that is incompatible with Cys binding because of a clash that would occur between Arg173 and the Cys α-amino group. This implies that *Re*CDOhom is not a CDO, clarifying a major question left unresolved in earlier studies. That *Re*CDOhom has over 50% sequence identity with a 3-mercaptopropionate dioxygenase provides a valuable clue to its specificity.

Future studies

Does CDO catalysis require a breathing motion? Is it prevented in the crystals? Given the amount of disruption that we see in the uncrosslinked active site, it is not surprising that no persulfenate forms. However, what is surprising is that the uncrosslinked enzyme is still active at all in solution. This difference raises in my mind the questions of whether the enzyme dynamics are possibly restricted in the

crystal in a way that alters its catalytic activity. One possibility is that CDO can "breathe" more in solution, such that even for the uncrosslinked C93A mutant, the enzyme adjusts so that Tyr could still participate in the catalytic triad. If such "breathing" motions of the enzyme were required during normal catalysis and this motion could not occur in the crystal due to crystal contacts, this could also help explain why persulfenate is captured in the active site in the crystal yet does not turn over to yield product.

Some evidence that a "breathing" motion does occur is seen in Figure 3.3, where the Cys93/Tyr157 pair, Arg60 and Tyr58 all shift slightly (~0.3 Å) and become more ordered in positions where they hydrogen bond with the persulfenate oxygen and α–amino group (Tyr157) or with its α-carboxylate (Arg60, Tyr58 and Tyr157). Additional difference map peaks distributed throughout the protein are associated with small global shifts in three lobes of the protein driven by the movement of the iron atom and the ligating His side chains. If the lobe including the iron ligands (residues 82-91, 100-124, 132-143, 161-179) is taken as a reference, the atoms at the edges of the other two lobes are shifted by ~1.0 Å (not shown).

These subtle shifts that occur upon substrate binding also indicate that the crystal is especially inhibiting the motion of the catalytic triad, as the Ser and His of the Ser-His-Tyr catalytic triad do not move as much as the Tyr. Future structure-function studies using nuclear magnetic resonance (NMR) in the presence and absence of substrate and analogs would be best suited to investigate such potential breathing motions of CDO.

Does crystalline CDO at higher pH values produce CSA? In addition to the possibility of the crystal contacts preventing the potentially necessary breathing motions for catalysis, another possibility to explain persulfenate capture at low pH values is related to the protonation state of Tyr157. If Cys-persulfenate is still able to form in the active site with protonated Tyr157, but that deprotonated Tyr157 is

required for subsequent steps in the catalytic cycle, then protonation states rather than dynamics may be the limiting factor.

At high pH values, Cys is captured in the active site and persulfenate is not observed, but it is not known if the enzyme is active or completely inactive such that it is unable to bind or react with oxygen at all. Recent pH dependence studies suggest that CDO is more active at these high pH values (5) and does not have substantial activity near pH=6. Using mass-spectrometry to try and detect CSA in the surrounding AML as was previously done at for soaks at low pH (4) would be the best approach to elucidating if the crystals are active at high pH.

It is possible that these studies will not show CSA production, in which case we would learn that a protonated Tyr is critical for stabilizing the Cys-persulfeante, but we would not learn if it was on-pathway or not because it still could be that the dynamics of the crystalline enzyme is limiting its activity.

How does Cys-Tyr crosslink formation depend on pH? One interesting aspect of CDO to consider is that the auto-catalyzed Cys-Tyr crosslink is forming through an energetically unfavorable intermediate. My reasoning is that crosslink formation proceeds through a mechanism that only happens in 1 out of every 800 turnovers (6). And a related possibility is that given arguments that Cys-persulfenate formation has also been said to be energetically unfavorable, might its formation be related to the chemistry that leads to crosslink formation.

To investigate these possibilities further, having a pH dependence of Cys-Tyr crosslink formation would be very useful. As methods have been developed for identifying both pure crosslinked (5) and non crosslinked (7) CDO, one could actually now use pure non crosslinked CDO and measure the rate of crosslink formation at a range of pH values. As one indication that the rate of crosslink formation could be different across different pH values, a recent study looked at the pH dependence of oxygen consumption (5). In this study, the rate of CSA formation was measured at a saturating Cys concentration (20 mM). The optimal pH values for oxygen

consumption (pH ~8.2) and CSA formation (pH ~8.6) were not identical. At high pH values (pH > 8.5), oxygen consumption was almost completely coupled with CSA production, and this coupling decreased with pH, with pH <7 having more than 25% oxygen consumed than CSA produced. It is possible that some of this excess oxygen consumption is associated with the chemistry that leads to Cys-Tyr crosslink formation and/or persulfenate formation rather than CSA production. While this is an interesting possibility, a study actually directly measuring the pH dependence of crosslink formation is needed before any conclusions can be made.

Does homocysteine outcompete Cys better in the non-crosslinked protein? Our structural studies suggest that Cys binds less tightly to non-crosslinked protein than to crosslinked protein, but that homocysteine binds similarily to both forms and that azide only binds to the crosslinked form. This shows that the crosslink does play an important role in inhibitor binding. Because cysteine and homocysteine have observable spectral changes when they bind to iron, obtaining binding constants to exclusively crosslinked or non crosslinked CDO would be quite doable.

What is the true pH dependence of the enzyme activity of CDO? Unfortunately, correlating our structurally-defined effective pKa values with those seen to influence the activity of CDO is difficult, because there has not been agreement on the pH dependence of activity. Based on the pH-profile by the Stipanuk group showing activity limited by apparent pKa values of 5.5 and 6.5 (6,8-9), we would infer that the deprotonation occurring in the crystal near pH=7.5 is blocking catalysis by preventing dioxygen binding or persulfenate formation. However, based on the pH-profile with maximal activity occurring between pH=7 and 9 (4,5,10), we would infer that the deprotonation event was generating the active form of the enzyme, either by allowing the persulfenate to complete the catalytic cycle or perhaps by through the formation of an alternative intermediate. Understanding why these assays differ and obtaining an accurate pH profile are important next steps in helping define the mechanism of CDO.

One possible issue with those assays which showed a pH optimum of ~6.1 was that they were carried out in the presence of added ferrous iron (6,8-9), and those studies which reported higher pH values did not have extra iron added (4,5,10). We considered the possibility that the added ferrous iron is altering the pH profile. One concern is that the decrease in activity at higher pH was not necessarily measuring the intrinsic CDO activity, but occurred because the Fe(II) added to the assay mix forms iron hydroxides and is hindered from getting into the CDO active site.

Preliminary data I collected also shows that non-enzymatic iron oxidation of Cys to CSA follows the same trend of pH dependence as observed in those assays that used added iron, with activity decreasing at pH > 6.5 (Figure 6.2). This preliminary evidence suggests that ferrous iron is influencing the pH dependence of activity, but further studies are needed to show the ferrous iron is influencing the relevance of the assay.

Two studies published after our work in Chapter 4 investigated the pH dependence of activity for rat CDO (5,11), neither of which added ferrous iron to the reaction mixture. Both of these studies showed a very similar pH dependence on CSA production (5). One issue with these kinetic characterizations on the pH dependence of activity is that the assays are carried out using a mixture of crosslinked to non crosslinked CDO. Our study (Chapter 4), suggests that the optimal pH for activity could be different between the crosslinked and non crosslinked populations, with several key active site pKa values changing. Using new methods to obtain homogenous populations of CDO, accurate pH profiles can now be obtained for the wild-type crosslinked and non-crosslinked enzyme. One technical difficulty of these studies would be the fact that non crosslinked would be converting to crosslinked during the course of the reaction; this could perhaps be overcome by short incubation times. Alternatively, the use of non-crosslinked variants (including bacterial CDOs) could substitute for the non-crosslinked enzyme, as these pH profiles have not been carried out yet either.

Would calculations with appropriate coordinates yield different results? Quantum mechanical calculations can greatly aid in elucidating reaction mechanisms, however using erroneous starting models, even if they are optimized to be more chemically reasonable, can conceivably have large effects on the calculated energetics. I suspect that part of the high energy calculate for the formation of the Cys-persulfenate is the necessary shift in the sulfur on the iron which must occur when using the erroneous structure that had been reported (12), but is minimal when the correct location that we observe in our Cys-bound structures is used. Another potentially influential flaw in those coordinates is the Arg60 position that remains in the incompatible position II (as discussed in Chapter 3). The only way to learn the impact of these coordinate errors will be to redo the calculations using the proper Cys-only coordinates from our high resolution structures.

Concluding remarks

I have presented here my research in structural biology that has advanced our understanding of the structure-function relations of the enzymes SsuE and CDO. I have described how these enzymes relate to sulfur metabolism and specifically how these studies have advanced the field. It is my hope that readers can appreciate the many ways that these structure-structure studies have improved our understanding these enzymes, and get a sense for how important structural approaches are for understanding the natural world.

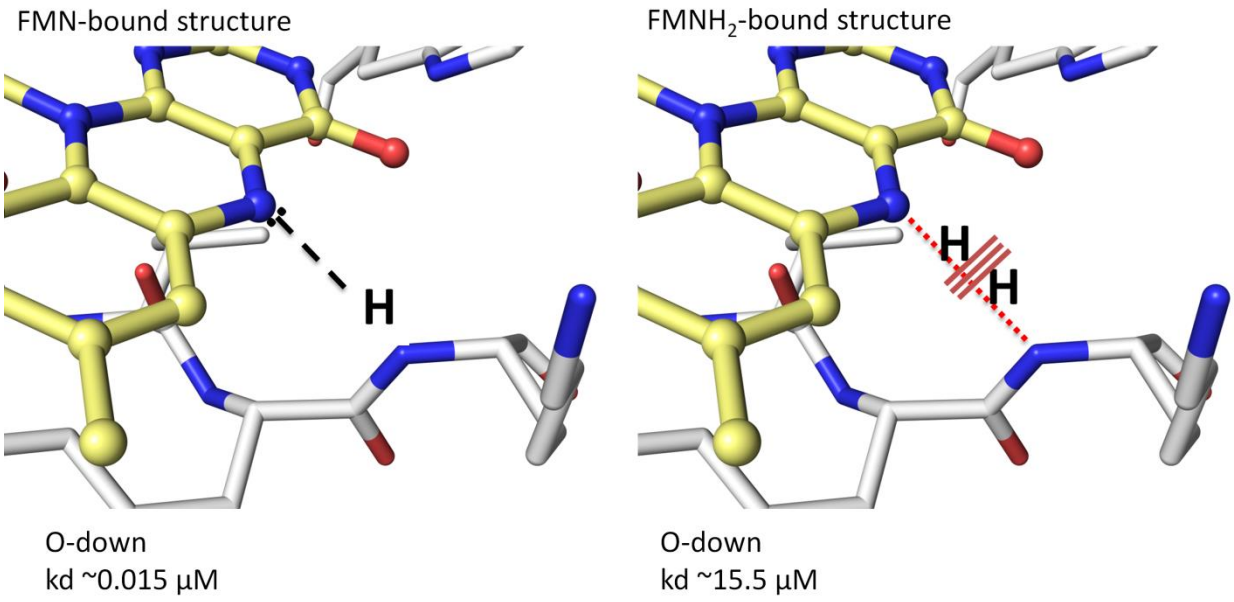


Figure 6.1. Proposal for the origin of the lower affinity for SsuE to $FMNH_2$. In flavodoxin (PDB entry 1fla) this peptide undergoes a peptide flip upon FMN reduction, in SsuE no such flip occurs. This alone is enough to explain the 1000-fold difference in binding affinity (4 kcal/mol) (13,14), since a H-bond is on the order of 1 to 3 kCal/mol and the clash between the hydrogens could account for the remaining difference.

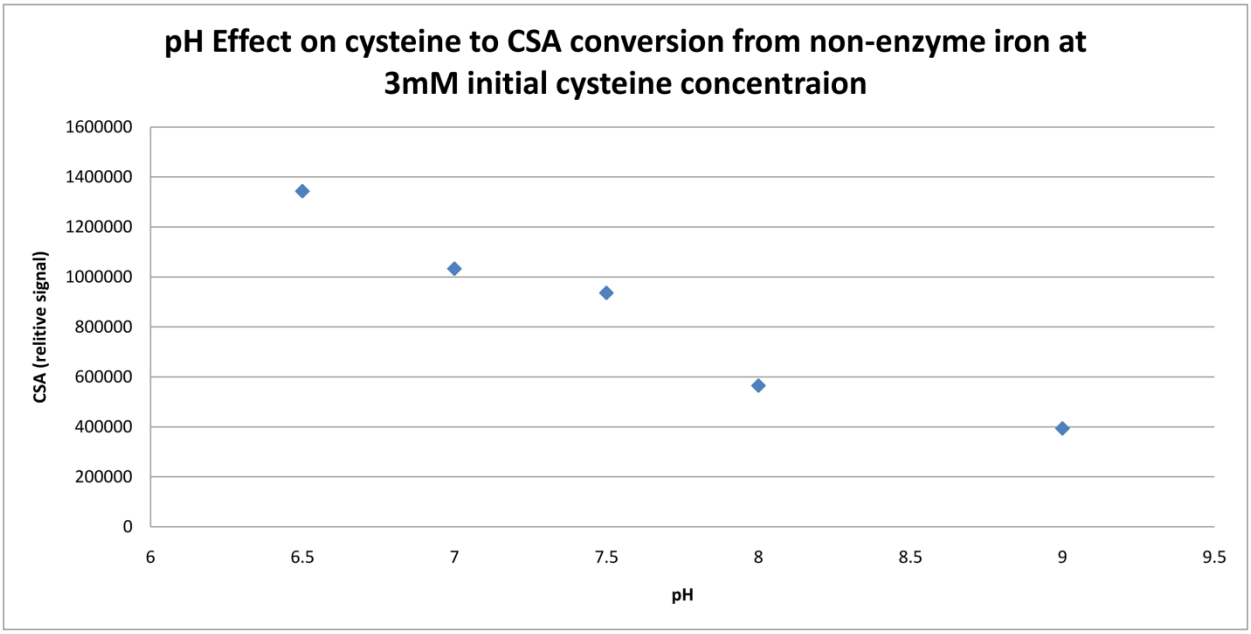


Figure 6.2. Preliminary results showing the pH Effect on Cys to CSA conversion from non-enzyme iron. In this assay 15 µL of a 0.5 mM Ferrous sulfate in 10 mM Tris 50 mM NaCl at pH 7.4 was mixed with a 385µL of a cysteine solution in a buffer composed of 100 mM MES, 100 mM Tris, and 100 mM acetate at the desired pH value. The final concentration of Cys in the reaction was 3 mM. The amount of CSA produced is about 10 times less than the enzymatic reaction which contains 15 µL of a 3.0 mg/mL solution of CDO.

References

1. Ellis, H. R. (2010) The FMN-dependent two-component monooxygenase systems. Arch Biochem Biophys 497, 1-12

2. Robbins, J. M., and Ellis, H. R. (2012) Identification of critical steps governing the two-component alkanesulfonate monooxygenase catalytic mechanism. Biochemistry 51, 6378-6387

3. Xiong, J., and Ellis, H. R. (2012) Deletional studies to investigate the functional role of a dynamic loop region of alkanesulfonate monooxygenase. Biochim Biophys Acta 1824, 898-906

4. Souness, R. J., Kleffmann, T., Tchesnokov, E. P., Wilbanks, S. M., Jameson, G. B., and Jameson, G. N. (2013) Mechanistic implications of persulfenate and persulfide binding in the active site of cysteine dioxygenase. Biochemistry 52, 7606-7617

5. Li, W., Blaesi, E. J., Pecore, M. D., Crowell, J. K., and Pierce, B. S. (2013) Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O_2 coupling efficiency in mouse cysteine dioxygenase. Biochemistry 52, 9104-9119

6. Dominy, J. E., Jr., Hwang, J., Guo, S., Hirschberger, L. L., Zhang, S., and Stipanuk, M. H. (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. J Biol Chem 283, 12188-12201

7. Njeri, C. W., and Ellis, H. R. (2014) Shifting Redox States of the Iron Center Partitions CDO between Crosslink Formation or Cysteine Oxidation. Arch Biochem Biophys DOI: 10.1016/j.abb.2014.06.001

8. Simmons, C. R., Hirschberger, L. L., Machi, M. S., and Stipanuk, M. H. (2006) Expression, purification, and kinetic characterization of recombinant rat cysteine dioxygenase, a non-heme metalloenzyme necessary for regulation of cellular cysteine levels. Protein Expr Purif 47, 74-81

9. Dominy, J. E., Jr., Simmons, C. R., Karplus, P. A., Gehring, A. M., and Stipanuk, M. H. (2006) Identification and characterization of bacterial cysteine

dioxygenases: a new route of cysteine degradation for eubacteria. J Bacteriol 188, 5561-5569

10. Chai, S. C., Jerkins, A. A., Banik, J. J., Shalev, I., Pinkham, J. L., Uden, P. C., and Maroney, M. J. (2005) Heterologous expression, purification, and characterization of recombinant rat cysteine dioxygenase. J Biol Chem 280, 9865-9869

11. Fellner, M., Doughty, L. M., Jameson, G. N., and Wilbanks, S. M. (2014) A chromogenic assay of substrate depletion by thiol dioxygenases. Anal Biochem 459C, 56-60

12. Gao, B., and Ellis, H. R. (2005) Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. Biochem Biophys Res Commun 331, 1137-1145

13. Zhan, X., Carpenter, R. A., and Ellis, H. R. (2008) Catalytic importance of the substrate binding order for the $FMNH_2$-dependent alkanesulfonate monooxygenase enzyme. Biochemistry 47, 2221-2230

Appendices

Appendix 1

Cysteine dioxygenase structures from pH 4 to 9: Consistent Cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH- supplemental information

Camden M. Driggers, Richard B. Cooley, Banumathi Sankaran, Lawrence L. Hirschberger, Martha H. Stipanuk, and P. Andrew Karplus

Published in *Journal of Molecular Biology* (2013), 425(17), 3121-3136.

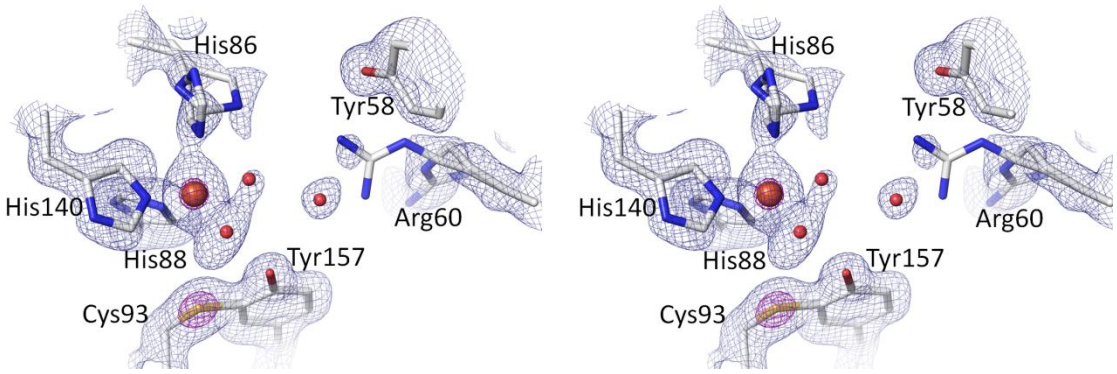


Figure A1.1. Partial iron loss in the pH=4.0 Cys structure. Stereo figure showing the active site density contoured at 1.2 ρ_{rms} (blue) and 5.0 ρ_{rms} (purple) shows weak complex density for the His86 side chain density that is weaker than that of the Cys93 Sγ, consistent with the iron having decreased occupancy. We modeled the occupancy at 80% by selecting occupancy that yields a B-factor that matches that expected based on a fully liganded structure. The refined model is shown including alternate conformations for the His86 and Arg60 side chains. Our interpretation is that the side chain of His86 becomes partially protonated at this pH, weakening iron ligation and allowing His86 to adopt an alternative conformation (shown).

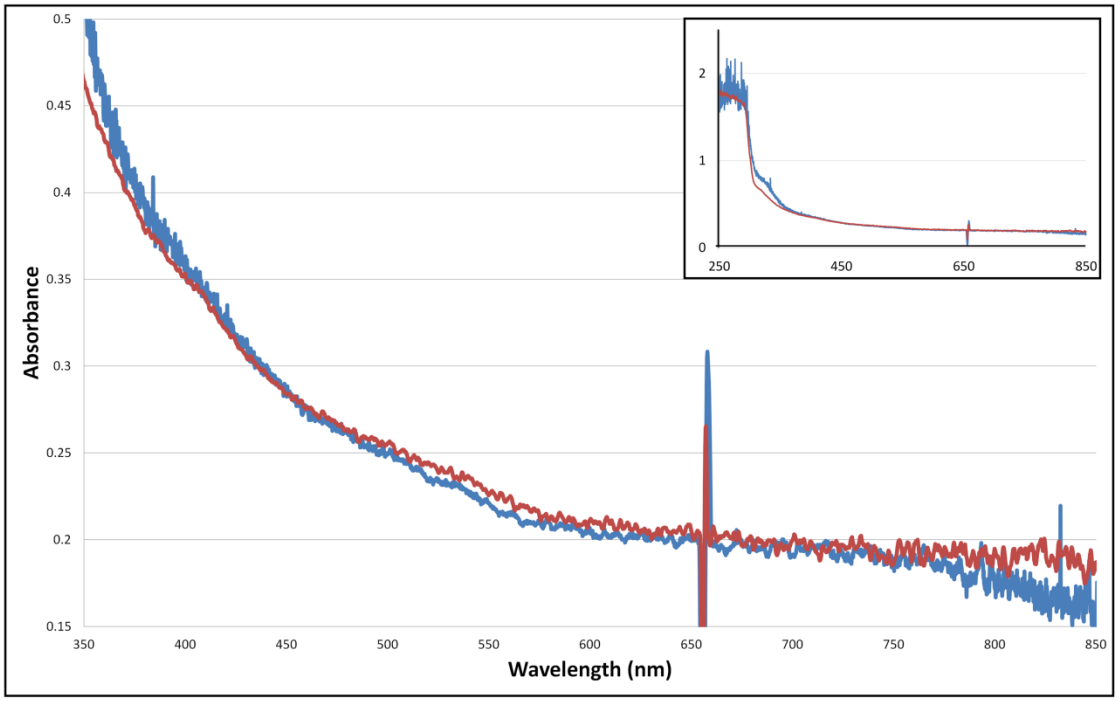


Figure A1.2. UV/VIS spectra of select CDO crystals before data collection. Shown are the spectra of a crystal containing a Cys-only complex (red trace; pH=8.0 + Cys soak) and a crystal containing a bound cys-persulfenate/persulfenic acid complex (blue trace; pH 6.2 + Cys soak). To facilitate comparisons, the blue trace shown has been origin shifted (by subtracting 0.207 absorbance units) from the original data so that its absorbance at 750 nm matches that of the red trace. The main panel shows no significant difference exists in the spectra surrounding the 640 nm region where an absorbance band would be expected for an Fe(III)-Cys complex. The inset shows the full spectral range collected. Diffraction data collected for each of these crystals after the spectra were taken confirmed the content of the structures. The one visible difference is a small additional absorbance that occurs for the CDO:Cys-persulfenate containing crystal at 325 nm. The spike at 650 nm is due to a technical aspect of how the spectra were collected. To test for dichroic absorption bands, spectra were taken for all crystal orientations around a 360˚ rotation axis with kappa 0 and using a kappa offset of 90 ˚ (i.e. a crystal tilt of 45 ˚). All the spectra were similar to the ones shown. Spectra were collected at National Synchrotron Light Source beamline X26C.