Structures of Arg- and Gln-type bacterial Cysteine Dioxygenase Homologs

Citation	Driggers, C. M., Hartman, S. J., & Karplus, P. A. (2015). Structures of Arg-and Gln-type bacterial cysteine dioxygenase homologs. Protein Science, 24(1), 154-161. doi:10.1002/pro.2587
DOI	10.1002/pro.2587
Publisher	John Wiley & Sons, Inc.
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Structures of Arg- and Gln-type bacterial Cysteine Dioxygenase Homologs

Camden M. Driggers[a], Steven J. Hartman[a], and P. Andrew Karplus[a*]

[a] Department of Biochemistry and Biophysics, 2011 Ag & Life Sciences Bldg, Oregon State University, Corvallis, OR 97331

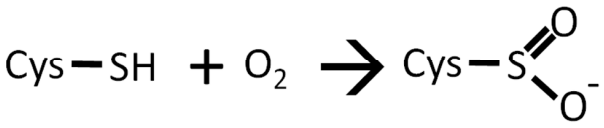
[*]Corresponding author: P. Andrew Karplus
Department of Biochemistry and Biophysics
2011 Ag & Life Sciences Bldg
Oregon State University, Corvallis, OR 97331
Phone: (541) 737-3200; Fax (541) 737-0481
E-mail: karplusp@science.oregonstate.edu

Abstract

In some bacteria, cysteine is converted to cysteine sulfinic acid by cysteine dioxygenases (CDO) that are only ~15-30% identical in sequence to 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). Here, we describe a structure from each of these enzyme types by analyzing 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 that is hydrogen bonding to an iron-bound diatomic molecule we have interpreted as dioxygen. Notably, the Arg position is not compatible with the mode of Cys binding seen in both rat CDO and *Bs*CDOs. As sequence alignments show that this newly discovered active site Arg is well conserved among "Gln-type" CDO enzymes, we conclude that the "Gln-type" CDO homologs are not authentic CDOs but will have substrate specificity more similar to 3-mercaptopropionate dioxygenases.

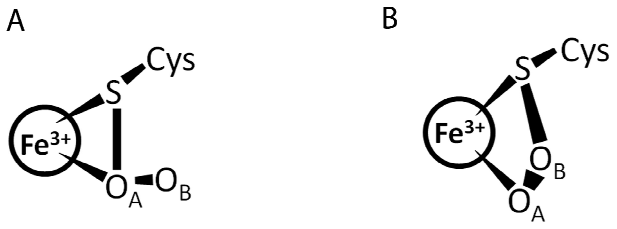
Introduction

Cysteine dioxygenase (CDO) is 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:



CDOs enzymatic mechanism is still not understood despite enzymatic,[1-5] spectroscopic,[6-9] and crystallographic[10-14] characterizations of wild type and mutant CDOs, studies with small molecule analogs of the metallocenter,[15,16] and quantum mechanical analyses of potential mechanisms.[17-20] All but one[5] of these studies have been done on rat/mouse (identical in sequence) or human CDO (99% identical in sequence to rat CDO). As reviewed recently,[21] for mammalian CDOs the ferrous iron is coordinated by three His and an ordered water molecule, and nearby is a key Tyr157 residue (rat CDO numbering) in a thioether crosslink with Cys93 and part of a Ser153-His155-Tyr157 catalytic triad. The thioether crosslink is autocatalyzed and increases activity 10-fold[1] or more.[3,22] Cys, upon binding, displaces the Fe-bound water and ligates the iron via its sulfhydryl and amine groups, with its carboxylate interacting with the side chains of Arg60, Tyr58 and the key catalytic Tyr157.[12]

Debate about the mechanism centers on whether Cys thiolate is first oxidized by the iron-proximal oxygen atom to form a persulfenate-type intermediate (seen in crystals and shown as structure “A” below) or by the iron distal oxygen atom to form a four-membered ring (shown as structure “B” below) that proceeds via O-O bond breakage to generate an FeIV-oxo intermediate.



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 enzyme forms that conserve mechanism but provide a distinct window into the chemistry.

Bacterial CDOs generally have sequence identities <30% compared with the mammalian enzymes[5] and have a Gly in place of Cys93, meaning that they have no Cys-Tyr crosslink. Also, some bacterial CDO homologs conserve Arg60 that interacts with Cys α-carboxylate and others have a Gln in its place;[5] we will call these "Arg-type" and "Gln-type" CDO homologs, respectively. Four "Arg-type" homologs were shown to have CDO activity,[5] but the only activity tests of a "Gln-type" homolog showed that the enzyme from *Variovorax paradoxus* had no CDO activity, but was active as a 3-mercaptopropionate dioxygenase.[23]

Structural studies of bacterial CDO homologs would be valuable for the field, and although they have not been described in the literature, two such structures have been solved by structural genomics efforts and been deposited in the Protein Data Bank[1]: PDB entry 3EQE (released 10/2008), solved by the NorthEast Structural Genomics Consortium (NESG), is an Arg-type CDO [with proven activity][5] from *Bacillus subtilis* (i.e. *Bs*CDO); and PDB entry 2GM6

[1] PDB entry 3USS (released January 2012; 2.7 Å resolution) is also a Gln-type bacterial CDO homolog (from *Pseudomonas aeruginosa* and 64% identical to *Re*CDOhom). It was deposited by the Jameson group that actively investigates CDOs; we do not describe it here as we presume they will be analyzing it.

(released 6/2006), solved by the Joint Center for Structural Genomics (JCSG), is a functionally uncharacterized *Ralstonia eutropha* CDO homolog (*Re*CDOhom). Notably, *Re*CDOhom is 55% identical in sequence to the *V. paradoxus* mercaptoproprionate dioxygenase. In the spirit of maximizing the value of these structures determined by structural genomics efforts, here we describe the *Bs*CDO and *Re*CDOhom structures and their comparisons with the mammalian CDO structures.

Results and Discussion

Compared with the original PDB entries, the analyses of *Bs*CDO crystals with Cys-bound that diffract to 2.3 Å resolution and extending the resolution of *Re*CDOhom from 1.84 to 1.65 Å have led to improved models with overall R_{free} values lowered by ~3 % for *Bs*CDO and ~1 % for *Re*CDOhom, even at the extended resolutions (Table 1). More importantly, we altered some key aspects of the models, most prominently placing a diatomic O_2-like molecule rather than a sulfate in the active site of *Re*CDOhom. Also, our work to reproduce the *Bs*CDO crystals led to the successful capture of a Cys bound in its active site. The remainder of the Results and Discussion section will first summarize broad features of the two proteins 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 1A). The *Bs*CDO structures have two copies in the asymmetric unit with the chains agreeing within ~0.5 Å. Chain A includes residues 1-154 and Chain B residues 1-

152 (of 161), and due to crystal contacts, residues 98-102 and the C-terminus of Chain A are more ordered. For *Re*CDOhom, there is one chain in the asymmetric unit and the modeled residues 11-202 all have reasonable electron density. As seen in a structure-based sequence alignment (Figure 1B), the sequence of *Bs*CDO is ~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 enough residues are present in the sequence to form it (Figure 1). One other notable secondary structural feature is that relative to rat and *Bs*CDO, *Re*CDOhom has a β-bulge insertion in the middle of strand β6 (Figure 1).

Active Sites

***Bs*CDO active site.** We will first describe the more informative 2.3 Å resolution *Bs*CDO structure with Cys-bound. Similar to known CDO structures,[11-14] the active site has a non-heme iron coordinated by three conserved residues His75, His77, His125 (Figure 2A). Also present 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. Interestingly, Ser137 of the catalytic triad does not receive a hydrogen bond, as the Trp77 donor present in mammalian CDO[13] is replaced with Ile66 in *Bs*CDO. Additional strong active site $2F_o$-F_c density present was well fit by Cys (Figure 2A). The bound Cys coordinates the iron in a bidentate fashion via its Sγ and N atoms, with the α-amino group location being defined stereochemically even though it is not well defined by electron density at this resolution (Figure 2A). The α-carboxylate 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 those in the Cys-complex (Figure 2B). As was done for PDB entry 3EQE, we modeled three peaks around the iron as waters, but found their refined B-factors at ~40 $Å^2$ were much lower than the ~70 $Å^2$ seen for the surrounding atoms. This implies that these peaks are actually not due to water. Because the density peaks match reasonably well with those seen in the 2.3 Å Cys complex, we attempted refining Cys in the active site and obtained reasonable B-factors and a clean difference map, supporting this assignment (see Figure 2B – semitransparent model). Nevertheless, the active site density at this resolution is not definitive, so to be conservative our final refinement and deposition used the minimal interpretation of three waters in the active site. To ensure that users of the coordinates are aware that Cys may be bound, we also include in the file, with occupancy set to zero, the coordinates and B-factors we obtained by refining a bound Cys at 100% occupancy.

***Re*CDOhom active site.** Unlike the four- and five-coordinate irons seen in ligand-free mammalian CDO at pH 6.2[11] and pH 8.0,[12] the *Re*CDOhom iron is six-coordinate and liganded by three histidines (His94, His96, His147), two water/hydroxide molecules, and a diatomic molecule we have modeled as dioxygen (Figure 2C). Evidence supporting the assignment as dioxygen are the 123˚ angle of approach to the iron, the acceptance by the iron-proximal atom of a hydrogen bond from Arg173, and the reasonable B-factors (see Fig. 2C legend) and lack of residual difference map features. Thus, we will consider it a dioxygen, even while recognizing it could be something else. Elsewhere in the active site, the Ser160-His162-Tyr164 catalytic triad and a Trp (Trp85) hydrogen bonding with Ser160 are conserved with mammalian CDO. 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 2C). Again, no Cys-Tyr crosslink is present.

As was noted in the introduction, the typical active site Arg is replaced with Gln67, and this side chain is only weakly ordered. In contrast, a novel ordered active site side-chain, that of Arg173, is sandwiched between the side chains of Ile168 and Phe185. In addition to interacting with dioxygen, Arg173 hydrogen bonds with Asp95-O, Ser187-OH, and a water molecule that in turn interacts with 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 one-residue deletion after strand β3 (at Thr97) and a two-residue insertion after strand β9 that makes a 3_{10}-helix (including Ile168) (Figure 1B). These two segments pack against each other to provide the environment of Arg173 (Figure 3A).

Comparison of *Bs*CDO, *Re*CDOhom and rat CDO active sites. An overlay of the active site of *Bs*CDO with Cys bound, *Re*CDOhom with dioxygen bound, and rat CDO with Cys-persulfenate bound shows the remarkably consistent placement of the iron and the residues equivalent to rat CDO Tyr157 and Tyr58 in all three structures (Figure 3A). Also visible in *Bs*CDO and rat CDO are their common positioning of the Cys and the residue equivalent to rat CDO Arg60. This provides strong evidence that *Bs*CDO and other bacterial CDOs conserving Arg60 will be mechanistically equivalent to rat CDO despite missing the Cys-Tyr crosslink.

Also of particular interest is what can be concluded about *Re*CDOhom and related enzymes having a Gln in the place of Arg60. The Gln67 side chain is located similarly to that of the Arg in *Bs*CDO, but cannot conserve all of its interactions with the Cys α-carboxylate. 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 the of oxygen binding inferred from CDO crystal structures (Figure 2D),[12-14] 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 complex with a productive substrate bound, we can already conclude that this enzyme *cannot* bind Cys in the same way as do the CDOs. This is because the well-fixed Arg173 guanidine group, as an obligatory hydrogen bond donor, would form an unfavorable clash with the hydrogen of a bound α-amino group (Figure 3A).

Arg173 as a key residue for Gln-type CDO homologs

The structure-based sequence alignment (Figure 1B) shows that *Re*CDOhom differs from both *Bs*CDO, and rat CDO, in having Arg173 (versus Met or Cys) and the short indels after strands β3 and β9 that give those loops unique backbone paths stabilizing the Arg173 sidechain. Strikingly, sequence conservation patterns show that both Arg173 and the indels are strongly conserved among Gln-type CDO homologs (Figure 3B), and are not present in the Arg-type enzymes (Figure 3C). This conservation pattern is consistent with a key functional role for Arg173 in the Gln-type CDO homologs, and leads us to conclude that none of the Gln-type enzymes are authentic CDOs – since they could not bind the Cys α-amino group in the expected way.

Taking into account the observation that the *V. paradoxus* 3-mercaptoproprionate dioxygenase[23] has 55% sequence identity with *Re*CDOhom, we propose that the Gln-type

enzymes are dioxygenases with specificity for a thiol substrates more similar to 3-mercaptoproprionate. At the same time we refrain from predicting the oxygen binding site in these enzymes, because the modeled oxygen binding site of *Re*CDOhom is distinct from that seen in CDO, but the key catalytic Tyr residues are similarly positioned (Figures 2D and 3A). This leads us to ask whether they actually have distinct ways of binding and activating oxygen, or whether one or both of the modes of oxygen binding seen are not relevant to catalysis. We suggest that the comparative study of authentic CDOs and these Gln-type dioxygenases will be very useful for resolving questions about mechanism; and we further suggest that given the amenability of *Re*CDOhom to high resolution structural studies, *Re*CDOhom itself would be an excellent system for pursuing further structural, kinetics, and spectroscopic studies of catalysis that would bring new insights into the CDO family of dioxygenases.

Making the Most of Structural Genomics Structures

The NIH funded Protein Structure Initiative invested heavily in structural genomics research centers with the dual goals of solving the structures of many representative proteins and protein domains as well as developing high throughput structure determination techniques.[24] These efforts were fruitful on both fronts and have accounted for ca. 13,000 PDB entries (as of July 2014). One unforeseen consequence of these efforts is the thousands of entries in the PDB that have not been described in the peer-reviewed literature. As is, these structures are of limited value to the broader scientific community because they will not show up in literature searches and because no expert having knowledge of both protein crystallography and the particular protein family has carefully vetted the structures for accuracy and information content. We suggest that, as exemplified by our work here on CDO homologs and elsewhere for two sets of

peroxiredoxin structures,[25,26] there now exists 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.

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* DH5α cells and used to transform *E. coli* BL21-DE3 chemically competent cells (Novagen). *Bs*CDO expression and purification basically followed the NESG protocols available in the PSI-knowledge base,[24] and yielded ~2 mg *Bs*CDO per liter culture that was stored frozen at ~10 mg/mL. Using crystallization conditions reported in the PDB entry, crystals grew in 2 to 7 d at 298 K in hanging drops of 4 uL *Bs*CDO stock and 4 µL reservoir containing 18% (w/v) PEG4000, 0.1 M potassium acetate, 0.05 M 2-(N-morpholino)ethanesulfonic acid at pH 6.0.

*Bs*CDO structure determination

Crystals were stored in an artificial mother liquor identical to the reservoir solution, and mounted by pulling through solutions having 20% glycerol as a cryoprotectant and plunging in liquid nitrogen. The Cys complex crystal was soaked 30 s with 100 mM cysteine at pH 7.0. Data for Cys-bound *Bs*CDO were collected at the Advanced Light Source beam-line 5.0.1 in a cryostream. Attempts to get higher resolution data for unliganded *Bs*CDO crystals were unsuccessful. Images were processed using Mosflm[27] and Aimless.[28] The high resolution cutoff

criterion was that the $CC_{1/2}$ statistic[29] be ~0.2 (Table 1). R_{free} flags were adopted from PDB entry 3EQE to 2.82 Å resolution, with a random 5% subset selected beyond.

Refinement at 2.3 Å resolution for 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[30] for manual model building, Molprobity[31] to monitor the model's stereochemical quality, and Phenix[32] for minimizations using one TLS group per chain. Sidechain rotomers were adjusted, Met1 was stubbed and an alternate Asn31 conformation was added in chain B, and 69 water molecules were added in places having $2F_o$-F_c electron density $\geq$1 ρ_{rms}, Fo-Fc density $\geq$3 ρ_{rms}, and reasonable potential hydrogen bonding. Only near the end was the bound Cys built. The final R/R_{free} was 19.0/26.2% (Table 1).

The unliganded *Bs*CDO refinement against the deposited data (PDB entry 3EQE) began from the refined Cys-complex model after Cys and active site waters were removed. Other waters were retained except four that shifted beyond 3.5 Å from the protein. The refinements quickly converged to R/Rfree of 17.7/25.4 at 2.82 Å resolution (Table 1), an improvement over the values of 24.1/29.6% recorded for PDB entry 3EQE. Evidence that this might also be a Cys complex is described in the results.

Polishing Refinement of *Re*CDOhom

Starting from the deposited coordinates and structure factors of *Re*CDOhom (PDB ID 2GM6), Phenix refinement using one TLS group and riding hydrogens, resulted in R/R_{free} of 17.3/21.2% at 1.84 Å resolution. During manual model building, additional waters, ethylene glycols, and a sulphate were built, and active site density that had been modeled as a sulfate was reinterpreted as a dioxygen. Alternate conformations were added for Glu38, Gly39 and 9 solvent

molecules and a Leu124 alternate conformation was removed. The R/R_{free} after this stage was 16.9/20.35%, slightly below the 18.2/20.7% values of the deposited entry 2GM6. Then original diffraction images provided by the JCSG were processed (as described above for *Bs*CDO images), yielding data to 1.65 Å resolution (Table 1). R_{free} flags were adopted from PDB entry 2GM6 to its limiting resolution of 1.84 Å, with a random 5% subset selected beyond. The extended resolution maps were slightly better defined, leading to a final model with further improved R/R_{free} values of 17.4/20.0 (Table 1) even at the extended resolution limit.

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

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.

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. 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

23. 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

24. 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

25. 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

26. 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

27. 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

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

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

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

31. 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

32. 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

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

38. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282-1288

Table 1. Data collection and Refinement Statistics for *Bs*CDO and *Re*CDOhom structures

	***Bs*CDO unsoaked** [a]	***Bs*CDO Cys-soak**	***Re*CDOhom**
Data collection			
Space group	$P4_322$	$P4_322$	
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.[29]

[d] Ramachandran statistics as defined by Molprobity.[31]

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

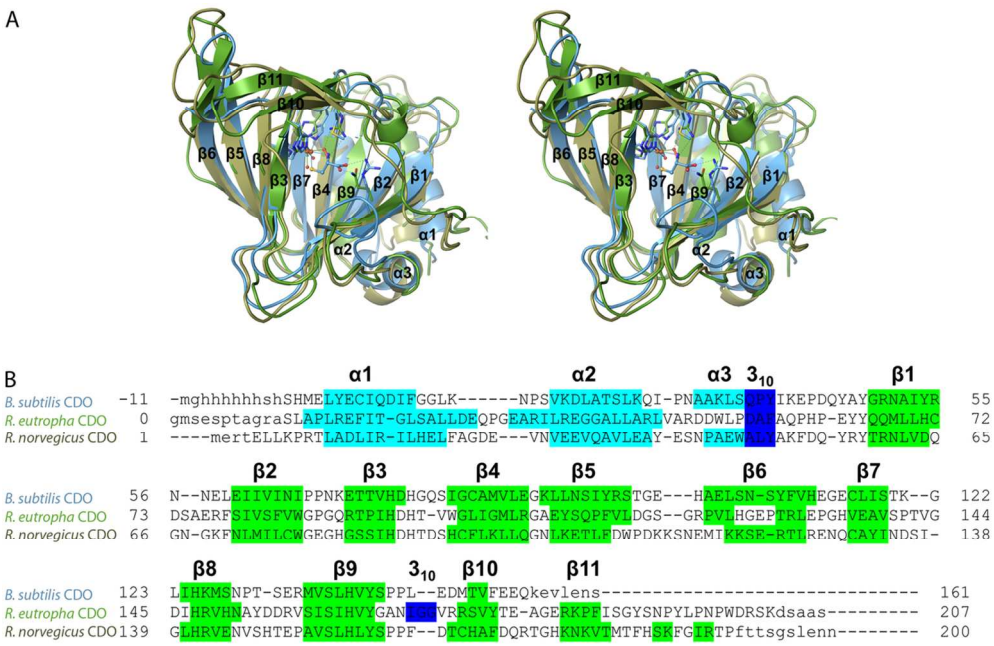
[f] Solvent in *Bs*CDO are waters, and in *Re*CDOhom are water, ethylene glycol and sulfate.

Figure 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 the similar overall structure of these three aligned homologs. 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 helices 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*CDOhom, and *R. norvegicus* CDO as generated using PROMALS,[35] and manually colored according to secondary structure as defined by DSSP.[36]

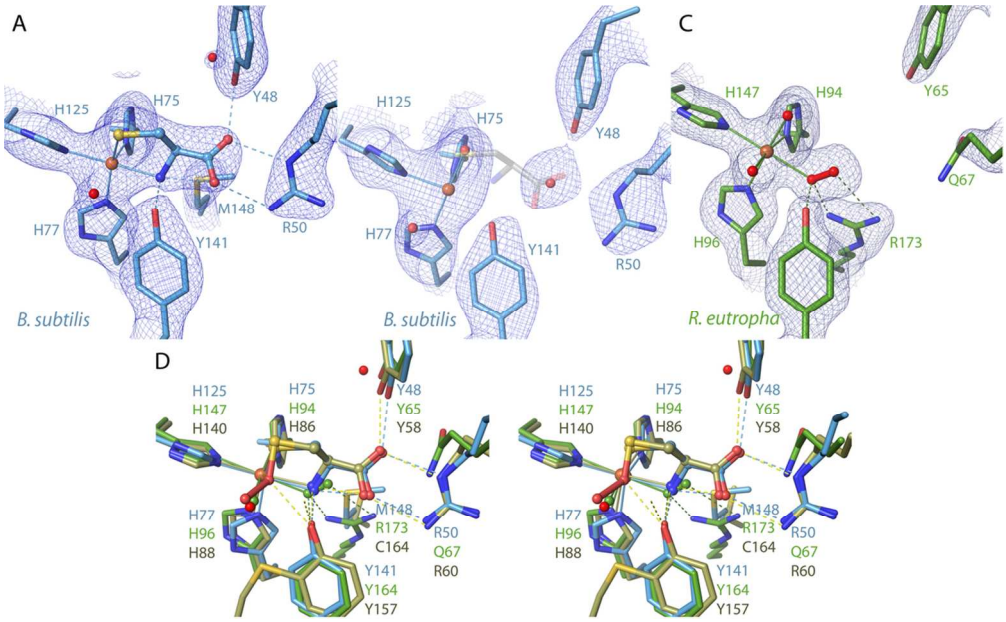
Figure 2. Active site structures of *Bs*CDO and *Re*CDOhom. Active site density for (A) *Bs*CDO at 2.3 Å resolution with cysteine-bound, (B) unsoaked *Bs*CDO at 2.8 Å resolution, and (C) *Re*CDOhom at 1.65 Å resolution with a diatomic molecule bound to the metal. All maps are $2F_o$-F_c electron density contoured at 1.2 ρ_{rms}. The putative dioxygen B-factors at ~40 $Å^2$ are comparable to the nearby Fe, Tyr164, Arg173 and water ligands which have B-factors in the 35-50 $Å^2$ range. (D) Local overlay of the active sites of *Bs*CDO (blue carbons), *Re*CDOhom (green carbons), and rat CDO (gold carbons).[12]

Figure 3. Arg173 packing interactions and associated residue conservation patterns. (A) Arg173 is well packed with Ile168 and Phe185 positioning the guanidine group with Asp95-O, Ser187-OH, Val175-O and Ser187-NH participating in hydrogen bonds (dashed lines). An overlaid rat CDO-Cys complex (semi-transparent) is shown with the clash between the α-amino group of

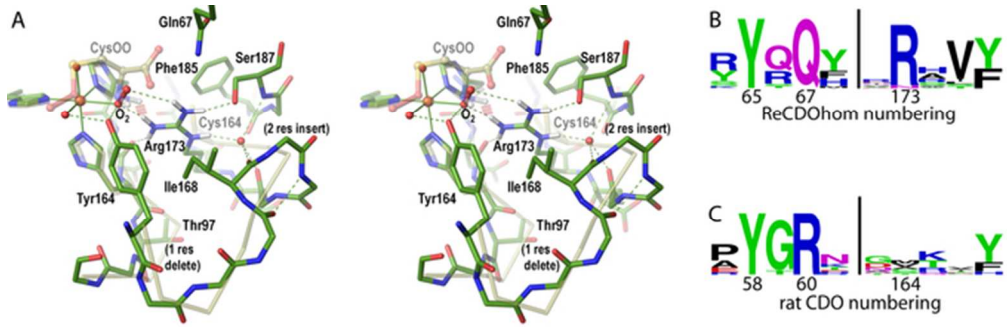
Cys-bound in the standard mode and Arg173 indicated (red bars). Similar packing interactions are also seen for the equivalent side chain in PDB entry 3USS, a lower resolution Gln-type CDO homolog structure also not yet described in the literature. (B) WebLogo[37] image of residue conservation pattern for Gln-type CDO homologs, The first segment contains the active site Tyr and Gln, and the second contains the conserved Arg173. (C) Same segments as panel B but for bacterial Arg-type CDO homologs. Aligned sequences included hits that contained key CDO active site residues and had E-value< 10^{-16} from BLASTP searches against the uniref50 database[38] obtained using *Re*CDOhom for panel B (20 sequences) and *Bs*CDO for panel C (13 sequences). The two sets of sequences were aligned together using PROSMALS-3D[35] before they were again separated for the Weblogo analysis.



Common cupin-fold of the bacterial CDO homologs. (A) Stereoview of the overlaid ribbon diagrams of R. norvegicus CDO (gold; PDB 4IEU)12, BsCDO (blue; ~2.3 Å Cα-rmsd vs. rat CDO) and ReCDOhom (green; ~2.2 Å Cα-rmsd vs. rat CDO) shows the similar overall structure of these three aligned homologs. The active site Arg/Gln and the new active site Arg of the ReCDOhom (Met in rat CDO) are shown along with the iron coordinating His residues and the cysteine substrate as bound to rat CDO and BsCDO. The secondary structure labels are shown, with all three alpha helices 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 BsCDO ReCDOhom, and R. norvegicus CDO as generated using PROMALS,35 and manually colored according to secondary structure as defined by DSSP.36
114x73mm (300 x 300 DPI)



Active site structures of BsCDO and ReCDOhom. Active site density for (A) BsCDO at 2.3 Å resolution with cysteine-bound, (B) unsoaked BsCDO at 2.8 Å resolution, and (C) ReCDOhom at 1.65 Å resolution with a diatomic molecule bound to the metal. All maps are 2Fo-Fc electron density contoured at 1.2 prms. The putative dioxygen B-factors at ~40 Å2 are comparable to the nearby Fe, Tyr164, Arg173 and water ligands which have B-factors in the 35-50 Å2 range. (D) Local overlay of the active sites of BsCDO (blue carbons), ReCDOhom (green carbons), and rat CDO (gold carbons).12
108x66mm (300 x 300 DPI)



Arg173 packing interactions and associated residue conservation patterns. (A) Arg173 is well packed with Ile168 and Phe185 positioning the guanidine group with Asp95-O, Ser187-OH, Val175-O and Ser187-NH participating in hydrogen bonds (dashed lines). An overlaid rat CDO-Cys complex (semi-transparent) is shown with the clash between the α-amino group of Cys-bound in the standard mode and Arg173 indicated (red bars). Similar packing interactions are also seen for the equivalent side chain in PDB entry 3USS, a lower resolution Gln-type CDO homolog structure also not yet described in the literature. (B) WebLogo37 image of residue conservation pattern for Gln-type CDO homologs, The first segment contains the active site Tyr and Gln, and the second contains the conserved Arg173. (C) Same segments as panel B but for bacterial Arg-type CDO homologs. Aligned sequences included hits that contained key CDO active site residues and had E-value< 10-16 from BLASTP searches against the uniref50 database38 obtained using ReCDOhom for panel B (20 sequences) and BsCDO for panel C (13 sequences). The two sets of sequences were aligned together using PROSMALS-3D35 before they were again separated for the Weblogo analysis.
57x18mm (300 x 300 DPI)